Draft Comparative Effectiveness Review

Number XX

Field Triage Guideline Revision: Glasgow Coma Scale: Systematic Review

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No.:
To Be Added for Final Version

Prepared by:
To Be Added for Final Version

Investigators:
To Be Added for Final Version

AHRQ Publication No. xx-EHCxxx
<Month Year>

Disclaimer

This draft report is based on research conducted by the EPC name to be added for final version under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. To Be Added for Final Version). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may periodically be assessed for the currency of conclusions. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at www.effectivehealthcare.ahrq.gov. Search on the title of the report.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

Suggested citation: To Be Added for Final Version

Preface

AHRQ, through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The National Highway Traffic Safety Administration (NHTSA) requested and provided funding for this report.

The reports and assessments provide organizations with comprehensive, evidence-based information on common medical conditions and new health care technologies and strategies. They also identify research gaps in the selected scientific area, identify methodological and scientific weaknesses, suggest research needs, and move the field forward through an unbiased, evidence-based assessment of the available literature. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review and public comment prior to their release as a final report.

AHRQ expects that the EPC evidence reports and technology assessments, when appropriate, will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

If you have comments on this evidence report, they may be sent by mail to the Task Order Officer (TOO) named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Andrew Bindman, M.D.
Director
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

David Niebuhr, MD, MPH, MSc
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Cathy Gotschall, Sc.D.
Senior Health Scientist
National Highway Traffic Safety Administration
U.S. Department of Transportation

Acknowledgments

To Be Added for Final Version

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who provided input to this report follows:

To Be Added for Final Version

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:

To Be Added for Final Version

Field Triage Guideline Revision: Glasgow Coma Scale: Systematic Review

Structured Abstract

Objectives. To assess the predictive utility, reliability, and ease of use of the total Glasgow Coma Scale (tGCS) versus the motor component of the Glasgow Coma Scale (mGCS) for field triage of trauma, as well as comparative effects on clinical decisionmaking and clinical outcomes.

Data Sources. MEDLINE®, CINAHL, PsycINFO, HAPI (Health & Psychosocial Instruments), and the Cochrane Databases (January 1995 through February, 2016). Additional studies were identified from reference lists and technical experts.

Study Selection. Studies on the predictive utility of the tGCS versus the mGCS or Simplified Motor Scale (SMS) (a simplified version of the mGCS), randomized trials and cohort studies on effects of the tGCS versus the mGCS on rates of over- or under-triage, and studies on interrater reliability and ease of use of the tGCS, mGCS, and/or SMS.

Data Extraction. One investigator abstracted details about study design, patient population, setting, screening method, followup, and results; a second investigator checked data for accuracy. Two investigators independently applied prespecified criteria to rate study quality. Discrepancies were resolved through consensus. Data on discrimination of tGCS versus mGCS and tGCS versus SMS were pooled using a random effects model.

Results. 32 studies met inclusion criteria; 24 studies addressed predictive utility and nine addressed interrater reliability or ease of use. No study assessed comparative effects on over- or under-triage or clinical outcomes. For in-hospital mortality, the tGCS is associated with slightly greater discrimination than the mGCS (pooled mean difference in area under the receiver operating characteristic [AUROC] 0.013, 95% confidence interval [CI] 0.007 to 0.019; I^2=59%, 11 studies; strength of evidence [SOE]: Moderate) or the SMS (pooled mean difference in AUROC 0.030, 95% CI 0.024 to 0.036, I^2=0%, 5 studies; SOE: Moderate). This means that for every 100 trauma patients, the tGCS is able to correctly discriminate 1 to 3 more cases of in-hospital mortality than the mGCS or the SMS. The tGCS is also associated with greater discrimination than the mGCS or SMS for receipt of neurosurgical interventions, severe brain injury, and emergency intubation (differences in AUROC ranged from 0.03 to 0.05; SOE: Moderate). Differences in discrimination between the mGCS versus the SMS were small (differences in the AUROC ranged from 0.000 to 0.01; SOE: Moderate).

Findings were robust in sensitivity and subgroup analyses based on age, type of trauma, study years, assessment setting (out-of-hospital versus emergency department), risk of bias assessment, and other factors. Differences between the tGCS, mGCS, and SMS in diagnostic accuracy (sensitivity, specificity) based on standard thresholds (≤15, ≤5, and ≤1) were small, based on limited evidence (SOE: Low). The interrater reliability of tGCS and mGCS appears to be high, but evidence was insufficient to determine if there were differences between scales (SOE: Insufficient). Three studies found the tGCS associated with a lower proportion of correct scores

than the mGCS (differences in proportion of correct scores ranged from 6% to 27%), though the difference was statistically significant in only one study (SOE: Low).

Limitations. Evidence on comparative predictive utility was primarily restricted to effects on discrimination. All studies on predictive utility were retrospective and the mGCS and SMS were taken from the tGCS rather than independently assessed. Most studies had methodological limitations. We restricted inclusion to English-language studies and were limited in our ability to assess publication bias. Studies on ease of use focused on scoring of video or written patient scenarios; field studies and studies on other measures of ease of use such as time required and assessor satisfaction were not available.

Conclusions. The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, and emergency intubation. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use. Research is needed to understand how use of the tGCS versus the mGCS or SMS impacts clinical outcomes and risk of over- or under-triage.

Contents

Tables

Figures

Appendixes

Executive Summary

Background

Unintentional injuries are the leading cause of death among people in the United States ages 1 to 44, and the third leading cause among people ages 45 to 64.[1] Among all age groups, motor vehicle crashes are the first or second leading cause of unintentional injury death.[2] In 2011, there were approximately 40,000,000 emergency department (ED) visits for injuries; of these approximately 2.5 million were due to trauma complications and unspecified injuries.[3] Approximately 18 percent of patients seen in the ED for an injury were transported by Emergency Medical Services (EMS) personnel.[4] Traumatic brain injury (TBI) is an important subset of trauma. Among an estimated 1.7 million annual cases of TBI, there are 52,000 deaths and 275,000 hospitalizations.[5] TBI is a contributing factor to about one third of injury related deaths in the United States. From 2001 to 2010, the rate of TBI-related ED visits increased from 421 to 716 per 100,000,[6] though the rate of deaths declined from 18.5 to 17.1 per 100,000 people.

Field Triage of Patients with Trauma

Field triage by EMS is a critical aspect of trauma systems, as it helps to identify seriously injured patients for transport to major trauma centers,[7] which have been shown to improve survival in such patients.[8,9] Appropriate decisions regarding transport are crucial because management of severely injured patients in a Level I or a Level II trauma center has been shown to be associated with improved clinical outcomes.[9] On the other hand, unnecessarily triaging patients to high level trauma care who do not require it may represent an inefficient use of staff and resources.[10]

EMS providers must rapidly triage individuals who have undergone trauma in challenging environments. Therefore, EMS providers must have assessment tools that are easy to use, reliable, and accurate. A key component of field triage for patients with suspected serious injury is level of consciousness assessment.[4] The Glasgow Coma Scale (GCS)[11,12] is an instrument widely used for assessment of consciousness at the site of injury, in emergency departments, and in hospitals, as well as to monitor progress or deterioration during treatment.[13] The GCS consists of three items (components): eye (scored 1 to 4), verbal (scored 1 to 5), and motor (scored 1 to 6). Scores on each of these components are added to obtain the total Glasgow Coma Scale (tGCS) score, ranging from 3 to 15. Lower scores on the tGCS indicate lower levels of consciousness, generally correlating with more severe injury associated with poorer prognosis and requiring more intensive care. For patients with TBI, scores of 3 to 8 are generally considered to denote severe head injury, 9 to 12 moderate, and 13 to 15 mild.[14] The 2011 field triage guidelines from the Centers for Disease Control and Prevention (CDC) National Expert Panel recommend transporting patients with tGCS scores of 13 or less to facilities providing the highest level of trauma care.[4]

In some circumstances (e.g., trauma victims who are intoxicated, intubated, or whose other injuries influence response) it may not be possible to accurately assess the verbal and eye components of the GCS. In these cases, assessments may be primarily based on the motor component of the Glasgow Coma Scale (mGCS) alone.[11,15-17] In addition, the mGCS has been proposed for assessment of trauma patients even when the tGCS can be obtained, since only one item is assessed, potentially increasing ease of use in the field.[18] mGCS scores of 5 or less are

considered an indication of patients with severe injury.[18,19] The Simplified Motor Score (SMS) has been proposed as a streamlined alternative to the mGCS; it is assessed on a three point scale (scored 0 to 2, with a score of 0 corresponding to 1 to 4 on the mGCS, 1 corresponding to 5 on the mGCS, and 2 corresponding to 6 on the mGCS).[20]

Decisions regarding the use of field triage instruments for level of consciousness should be based on how they perform in comparison to the tGCS. The ultimate goal of selecting one risk prediction instrument over another is to improve clinical outcomes (e.g., mortality). However, information on clinical outcomes is often lacking and decisions about their use must often be based on how they perform on intermediate outcomes. Intermediate outcomes include measures of over- or under-triage (i.e., the degree to which patients are unnecessarily transported to a Level I or II trauma center [over-triage] or not transported to a Level I or II trauma center [under-triage]) or predictive utility, as assessed using measures of discrimination (ability of an instrument to distinguish patients with the disease from those without), calibration (how well predicted risk correlates with actual risk), standard measures of diagnostic accuracy (e.g., sensitivity, specificity, and predictive values), or adjusted risk estimates (e.g., odds ratio, relative risk, hazards ratio).[21] Other factors that could inform selection of field triage risk assessment instruments include intra- and interrater reliability and ease of use (e.g., time to administer the instrument and amount of missing data).[12,22,23]

A number of factors could impact the performance of field assessment instruments. These include variability in patient populations (e.g., type of trauma, demographic characteristics, presence and severity of intoxication, and medical comorbidities), level of training and certification of administering personnel (e.g., Emergency Medical Technician [EMT],[24] EMT-Intermediate, Advanced EMT/Paramedic, physician, or nurse[25]), receipt of field interventions (e.g., medications, intubation), setting (e.g., country, urban vs. rural) or timing of assessment relative to injury occurrence. Evidence about field triage instruments frequently relies on extrapolation from studies conducted in EDs, as this environment is more controlled and easier to study.[26] However, the performance of the tGCS and mGCS may be different when administered soon after injury by EMS personnel in the field as opposed to later by ED personnel, after destination decisions have already been made and patients have been stabilized with initial interventions.

During the development of field triage guidelines and algorithms by the CDC National Expert Panel in 2011,[4] use of the mGCS was considered a way to potentially simplify field triage. However, the mGCS was not adopted due in part to lack of evidence about the comparative accuracy and reliability of the mGCS relative to the tGCS. However, more evidence is now available on the mGCS.

Scope of Review and Key Questions

The research questions were initially developed by the National Highway Traffic Safety Administration and revised with input from a Technical Expert Panel. The Key Questions focus on predictive utility, over- and under-triage, clinical outcomes of the tGCS versus the mGCS or the Simplified Motor Score (SMS), as well as reliability and ease of use. We included studies of children and adults with known or suspected trauma, with assessment using the tGCS, the mGCS, or the SMS. For studies evaluating measures of diagnostic accuracy (sensitivity, specificity, predictive values), we focused on studies that used standard cutoff scores (≤13 for tGCS and ≤5 for mGCS), but also included studies that used alternative cutoffs or modifications of the tGCS and mGCS. For all Key Questions, we included cohort studies and randomized trials

that directly compared the tGCS versus the mGCS or SMS. For Key Question 4 (reliability and ease of use), we also included cross-sectional studies and studies that assessed one of these scales and for Key Question 1a (predictive utility) we included studies that assessed one of these scales if they addressed one of the subpopulations specified in the Key Questions not addressed well in the head-to-head studies.

For Key Question 1, we included measures of predictive utility for mortality, morbidity, markers of severe injury or utilization indicators of severe injury, as measured by diagnostic accuracy, adjusted risk estimates, measures of discrimination (e.g., the c-index), measures of calibration (e.g., the Hosmer-Lemeshow test), or risk reclassification rates. For Key Question 2, we included studies that reported the proportion of patients who were over- or under-triaged (e.g., the proportion transferred to a higher or lower level of care), for Key Question 3, we included studies that reported clinical outcomes, and for Key Question 4, we included outcomes that assessed reliability (e.g., interrater and intrarater kappa) or ease of use (e.g., time to complete, measures of missing data, user reported satisfaction).

For all Key Questions we included prospective and retrospective studies in which the tGCS, mGCS, or SMS was administered soon after injury (conducted in the field/out-of-hospital setting by EMS personnel) or immediately upon arrival to the ED, or that were based on trauma registry data collected in the field or in the ED.

The report focuses on the following Key Questions:

Key Question 1. In patients with known or suspected trauma, what is the predictive utility of the tGCS compared with the mGCS score for predicting mortality, morbidity, Injury Severity Score (ISS) of 16 or greater, head Abbreviated Injury Scale (AIS) score greater than 2 or greater than 3, presence of intracranial hemorrhage, and utilization indicators of severe injury (e.g., receipt of intracranial monitoring within 48 hours of admission, receipt of surgery within 12 hours of admission, or early intubation [in the field or immediately upon presentation to the ED])?

Key Question 1a. How does predictive utility vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medications in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 2. In patients with known or suspected trauma, what are the comparative effects of the tGCS compared with the mGCS on over- and under-triage (e.g., proportion of patients misclassified with regard to measures of injury severity or need for early interventions and transport to a lower vs. higher level of care)?

Key Question 2a. How do effects on clinical decisionmaking vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 3. In patients with known or suspected trauma, what is the comparative effectiveness of the tGCS compared with the mGCS on clinical outcomes (e.g., mortality, morbidity, quality of life)?

Key Question 3a. How do effects on clinical outcomes vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 4. In patients with known or suspected trauma, what is the comparative reliability (e.g., interrater and intra-rater kappa) and ease of use (e.g., time to complete, amount of missing data, user reported satisfaction) of the tGCS compared with the mGCS score?

Key Question 4a. How do comparative reliability and ease of use vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Methods

Literature Search Strategy

This review includes studies published since January, 1995. This search start date was selected because of changes in trauma care over time; only five states had fully implemented trauma systems in the early 1990s.[27] In addition, the first studies to compare the predictive utility of the mGCS versus the tGCS were published in 1998 and 2003.[18,19]

The Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, CINAHL, PsycINFO, HAPI (Health & Psychosocial Instruments) and Ovid MEDLINE (January, 1995 through February, 2016) were searched for relevant studies and systematic reviews. Investigators also manually reviewed reference lists of relevant studies and searched for unpublished studies in ClinicalTrials.gov.

Risk of Bias Assessment of Individual Studies

A single investigator abstracted details about study design, patient population, comparison groups, setting, screening method, analysis, followup, and results. A second investigator reviewed data abstraction for accuracy. By using prespecified criteria for risk prediction studies and cross-sectional studies, two investigators independently rated the quality of studies (good, fair, poor) and resolved discrepancies by consensus.

Data Synthesis

We applied a “best evidence” approach in which higher quality evidence (based on study design, risk of bias, and use of head-to-head vs. indirect comparisons) is prioritized. We did not exclude studies rated high risk of bias a priori, but performed sensitivity analyses to determine how their exclusion would impact conclusions. Within each Key Question, we qualitatively synthesized overall findings and assessed how potential modifiers of effects (e.g. patient characteristics, characteristics of the people administering the instrument, threshold used for the tGCS or mGCS, timing, or setting) impacted results, as well as study design characteristics (type of study, risk of bias). We performed meta-analysis using random effects models, using the

DerSimonian-Laird model with Stata/IC 13.1 (StataCorp LP, College Station, TX), based on similarities in the populations, interventions, comparisons, and settings evaluated. We also performed analyses using the Profile Likelihood method. Stratified and sensitivity analyses were performed on the potential modifiers of effects.

We evaluated any differences in conclusions based on direct versus indirect comparisons, as assessments of comparative diagnostic accuracy based on direct comparisons can differ from those based on indirect comparisons, and did not combine direct and indirect evidence.

Strength of the Body of Evidence

For all comparisons and outcomes we assessed the strength of evidence using the approach described in the AHRQ Methods Guide, based on the overall risk of bias (graded low, moderate, or high); the consistency of results across studies (graded consistent, inconsistent, or unable to determine when only one study was available); the directness of the evidence linking the intervention and health outcomes (graded direct or indirect); the precision of the estimate of effect, based on the number and size of studies and confidence intervals [CIs] for the estimates (graded precise or imprecise); and reporting bias (suspected of undetected). Assessments of reporting bias were based on whether studies defined and reported primary outcomes and whether we identified relevant unpublished studies.

Results

Results of Literature Searches

Database searches resulted in 4,306 potentially relevant citations. After dual review of abstracts and titles, 690 articles were selected for full-text review. After dual review of full-text articles, 32 studies were included.

Key Question 1. Predictive Utility

Twenty-four studies evaluated predictive utility.[17-20,28-47] Differences between the tGCS, mGCS, and SMS in discrimination (area under the receiver operating characteristic [AUROC]) for in-hospital mortality, neurosurgical intervention, severe brain injury, and emergency intubation were <0.05. Results were similar in subgroups stratified by age (child vs. mixed populations of adults and children), type of trauma (TBI vs. mixed trauma), field versus ED assessment, and other subgroup and sensitivity analyses. Main findings are summarized in Table ES-1 and below.

- In-hospital mortality
  - For the tGCS versus the mGCS, the pooled AUROC was 0.886 (95% CI 0.863 to 0.908) versus 0.864 (95% CI 0.839 to 0.890), respectively with a pooled mean difference of 0.013 (95% CI 0.007 to 0.019; I^2=59%), based on 11 studies (strength of evidence [SOE]: Moderate).
  - For the tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5), differences in sensitivity ranged from 0 percent to 3 percent; difference in specificity ranged from 0 percent to 5 percent in favor of the mGCS, though the CIs overlapped in each study (SOE: Low).

- For the tGCS versus the SMS, the pooled AUROC was 0.884 (95% CI 0.852 to 0.916) versus 0.840 (95% CI 0.802 to 0.878), respectively, for a mean difference of 0.030 (95% CI 0.024 to 0.036, I^2=0%), based on five studies (SOE: Moderate).
- One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity versus the SMS (cutoff of ≤1) (75%, 95% CI 73 to 76 vs. 72%, 95% CI 70 to 74) and slightly lower specificity (88%, 95% CI 87 to 88 vs. 89%, 95% CI 89 to 87) (SOE: Low).
- For the mGCS versus the SMS, the mean difference in the pooled AUROC was 0.014 (95% CI 0.006 to 0.021, I^2=0%), based on four studies (SOE: Moderate).

- Neurosurgical intervention
  - For the tGCS versus the mGCS, the pooled AUROC was 0.798 (95% CI 0.754 to 0.842) versus 0.769 (95% CI 0.722 to 0.815), respectively, with a mean difference of 0.027 (95% CI 0.020 to 0.034; I^2=0%), based on six studies (SOE: Moderate).
  - One study found no clear differences between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in accuracy for identifying people undergoing craniotomy (sensitivity 63%, 95% CI 38 to 84 vs. 68%, 95% CI 43 to 87; and specificity 82%, 95% CI 80 to 84 vs. 83%, 95% CI 81 to 85) (SOE: Low).
  - For the tGCS versus the SMS, the pooled AUROC was 0.809 (95% CI 0.766 to 0.853) versus 0.769 (95% CI 0.711 to 0.827), respectively, with a mean difference of 0.032 (95% CI 0.025 to 0.039, I^2=0%), based on five studies (SOE: Moderate).
  - One study found the out-of-hospital tGCS (cutoff of ≤13) associated with higher sensitivity than the SMS (cutoff of ≤1) for identifying patients who underwent neurosurgical intervention (60%, 95% CI 56 to 63 vs. 53%, 95% CI 49 to 56) and slightly lower specificity (85%, 95% CI 84 to 85 vs. 86%, 95% CI 86 to 87) (SOE: Low).
  - For the mGCS versus the SMS, the mean difference in the pooled AUROC was 0.002 (95% CI -0.005 to 0.010, I^2=0%), based on four studies (SOE: Moderate).
- Severe brain injury
  - For the tGCS versus the mGCS, the pooled AUROC was 0.791 (95% CI 0.734 to 0.827) versus 0.720 (95% CI 0.666 to 0.774), respectively, with a mean difference of 0.050 (95% CI 0.034 to 0.065; I^2=57%), based on five studies (SOE: Moderate).
  - One study found no difference between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in sensitivity (62%, 95% CI 55 to 68 vs. 61%, 95% CI 54 to 67) or specificity (85%, 95% CI 83 to 88 vs. 89%, 95% CI 88 to 91) for identifying people with severe head injury (defined as head AIS score of ≥4) (SOE: Low).
  - For the tGCS versus the SMS, the pooled AUROC was 0.763 (95% CI 0.710 to 0.815) versus 0.713 (95% CI 0.654 to 0.771), respectively, with a mean difference of 0.048 (95% CI 0.038 to 0.059, I^2=72%), based on five studies (SOE: Moderate).
  - One study found out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for severe brain injury based on presence of head CT imaging findings (45%, 95% CI 44 to 46 vs. 41%, 95% CI 40 to 42) and similar specificity (89%, 95% CI 89 to 90 vs. 90%, 95% CI 90 to 91) (SOE: Low).
  - For the mGCS versus the SMS, there was no difference in the AUROC (mean difference 0.000, 95% CI -0.008 to 0.007, I^2=0%), based on four studies (SOE: Moderate).

- Emergency intubation
  - For the tGCS versus the mGCS, the pooled AUROC was 0.851 (95% CI 0.794 to 0.908) versus 0.807 (95% CI 0.735 to 0.880), respectively, with a mean difference of 0.038 (95% CI 0.023 to 0.052; I^2=72%), based on five studies (SOE: Moderate).
  - For the tGCS versus the SMS, the pooled AUROC was 0.843 (95% CI 0.823 to 0.864) versus 0.783 (95% CI 0.747 to 0.819), respectively, with a mean difference of 0.040 (95% CI 0.030 to 0.050, I^2=55%), based on five studies (SOE: Moderate).
  - One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for identifying people who underwent emergency intubation (76%, 95% CI 74 to 77 vs. 73%, 95% CI 71 to 74) and slightly lower specificity (89%, 95% CI 89 to 89% vs. 91%, 95% CI 90 to 91) (SOE: Low).
  - For the mGCS versus the SMS, there was no difference in the AUROC (mean difference 0.000, 95% CI -0.007 to 0.007, I^2=0%), based on four studies (SOE: Moderate).
- Trauma center need
  - One study that utilized National Trauma Data Bank (NTDB) data (n=811,143) found small differences between the tGCS versus the mGCS in the AUROC (0.62 vs. 0.61), sensitivity (30% vs. 27%), and specificity (93% vs. 95%) for need of trauma center care (defined as ISS score of >15, intensive care unit [ICU] admission >24 hours, need for urgent surgery, or death in the ED) (SOE: Low).
- Severe injury
  - One study (n=104,035) of children with TBI in the NTDB found the tGCS was better able to discriminate those with major injury (defined as an ISS score of >15) from those without major injury (AUROC 0.720, 95% CI 0.715 to 0.724 vs. 0.681, 95% CI 0.677 to 0.686) (SOE: Low).
- Age: Effects on discrimination between the tGCS versus the mGCS were similar in studies that enrolled children and those that enrolled mixed populations of adults and children (SOE: Low).
- Type of trauma: Effects on discrimination between the tGCS versus the mGCS were similar in studies that evaluated patients with TBI and those that enrolled mixed trauma patients (SOE: Low).
- Out-of-hospital versus ED assessment: One study of adults found no differences between out-of-hospital and ED GCS scores on discrimination for mortality or neurosurgical intervention but another study of adults or children found out-of-hospital GCS scores associated with higher discrimination for mortality than ED scores (AUROC 0.754 vs. 0.635, p not reported). Effects on discrimination between the tGCS versus the mGCS were similar in studies that evaluated out-of-hospital GCS scores and those that used ED scores (SOE: Insufficient).
- No study evaluated how intoxication status, blood pressure, intubation status, receipt of field intubation, or level/training of field assessors impacts comparative predictive utility of the tGCS versus the mGCS or SMS.

Table ES-1. Pooled AUROC results of head-to-head studies

Outcome and analysis	tGCS vs. mGCS, difference in AUROC (95% CI)	Number of studies	I^2	tGCS vs. SMS, difference in AUROC (95% CI)	Number of studies	I^2
Mortality, overall	0.013 (0.007 to 0.019)	11	59%	0.030 (0.024 to 0.036)	5	0%
Adults or mixed	0.017 (0.015 to 0.020)	9	0%	0.030 (0.024 to 0.036)	5	0%
Children	0.006 (0.002 to 0.011)	2	0%	--	--	--
Excluding NTDB studies	0.014 (0.008 to 0.019)	9	0%	0.030 (0.024 to 0.036)	5	0%
Excluding studies with potential overlap*	0.013 (0.005 to 0.020)	8	68%	0.031 (0.023 to 0.039)	3	0%
Out-of-hospital GCS	0.012 (0.004 to 0.020)	6	74%	0.031 (0.023 to 0.039)	3	0%
ED GCS	0.020 (0.006 to 0.034)	3	20%	0.030 (0.020 to 0.039)	2	0%
U.S. setting	0.013 (0.007 to 0.019)	9	63%	0.030 (0.024 to 0.036)	5	0%
TBI patients	0.009 (-0.002 to 0.020)	3	0%	--	--	--
Low risk of bias studies	0.017 (0.015 to 0.020)	5	0%	0.030 (0.022 to 0.037)	3	0%
Enrollment before 2006	0.016 (0.012 to 0.020)	9	11%	0.030 (0.024 to 0.036)	5	0%
Enrollment after 2006	0.006 (0.001 to 0.011)	2	0%	--	--	--
Neurosurgical intervention, overall	0.027 (0.020 to 0.034)	6	0%	0.032 (0.025 to 0.039)	5	0%
Adults or mixed	0.026 (0.019 to 0.034)	5	0%	0.032 (0.025 to 0.039)	5	0%
Children	0.034 (0.009 to 0.059)	1	--	--	--	--
Excluding studies with potential overlap*	0.021 (0.008 to 0.034)	3	0%	0.038 (0.024 to 0.052)	3	19%
Out-of-hospital GCS	0.021 (0.008 to 0.034)	3	0%	0.038 (0.024 to 0.052)	3	19%
ED GCS	0.030 (0.020 to 0.039)	2	0%	0.029 (0.020 to 0.038)	2	0%
U.S. setting	0.027 (0.020 to 0.034)	6	0%	0.032 (0.025 to 0.039)	5	0%
TBI patients	0.017 (-0.022 to 0.056)	2	66%	--	--	--
Low risk of bias studies	0.026 (0.019 to 0.034)	4	0%	0.029 (0.021 to 0.037)	3	0%
Enrollment before 2006	0.028 (0.020 to 0.035)	5	0%	0.032 (0.025 to 0.039)	5	0%
Enrollment after 2006	0.019 (-0.009 to 0.047)	1	--	--	--	--
Severe brain injury, overall	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Adults or mixed	0.046 (0.038 to 0.054)	4	0%	0.048 (0.038 to 0.059)	5	72%
Children	0.121 (0.068 to 0.174)	1	--	--	--	--
Excluding NTDB studies	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Excluding studies with potential overlap*	0.065 (0.020 to 0.111)	3	76%	0.051 (0.034 to 0.068)	3	74%
Out-of-hospital GCS	0.041 (0.028 to 0.053)	2	0%	0.051 (0.034 to 0.068)	3	74%
ED GCS	0.060 (0.028 to 0.093)	3	73%	0.044 (0.030 to 0.059)	2	51%
U.S. setting	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%

Outcome and analysis	tGCS vs. mGCS, difference in AUROC (95% CI)	Number of studies	I^2	tGCS vs. SMS, difference in AUROC (95% CI)	Number of studies	I^2
TBI patients	--	--	--	--	--	--
Low risk of bias studies	0.046 (0.038 to 0.053)	3	0%	0.044 (0.035 to 0.053)	3	25%
Enrollment before 2006	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Enrollment after 2006	--	--	--	--	--	--
Emergency intubation, overall	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
Adults or mixed	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
Children	--	--	--	--	--	--
Excluding studies with potential overlap*	0.031 (0.012 to 0.050	3	65%	0.033 (0.025 to 0.040)	3	0%
Out-of-hospital GCS	0.031 (0.012 to 0.050)	3	65%	0.033 (0.025 to 0.040)	3	0%
ED GCS	0.048 (0.039 to 0.058)	2	0%	0.048 (0.039 to 0.057)	2	0%
U.S. setting	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
TBI patients	0.011 (-0.010 to 0.032)	1	--	--	--	--
Low risk of bias studies	0.037 (0.022 to 0.052)	4	79%	0.046 (0.038 to 0.054)	3	0%
Enrollment before 2006	0.046 (0.038 to 0.053)	4	0%	0.040 (0.030 to 0.050)	5	55%
Enrollment after 2006	0.018 (0.005 to 0.031)	1	--	--	--	--

AUROC=area under the receiver operating characteristics curve; CI=confidence interval; ED= emergency department; GCS= Glasgow Coma Scale; mGCS= motor Glasgow Coma Scale; NTDB= National Trauma Data Bank; SMS= Simplified Motor Score; TBI=traumatic brain injury; tGCS= total Glasgow Coma Scale; U.S.= United States of America

*When multiple studies published from the same trauma center, analysis restricted to the most recent study using out-of-hospital GCS scores (excluded Gill 2005,[20] Haukoos 2007,[38] Acker 2014[28])

Key Question 2. Over- and Under-Triage Rates

No study evaluated comparative effects of the tGCS versus the mGCS or SMS on over- or under-triage rates.

Key Question 3. Effectiveness of Clinical Outcomes

No study evaluated comparative effects of the tGCS versus the mGCS or SMS on clinical outcomes.

Key Question 4. Interrater Reliability and Ease of Use

Ten studies evaluated interrater reliability or ease of use.[39,43,48-55] Evidence on comparative interrater reliability and ease of use was very limited. There were few head-to-head studies, studies had methodological limitations, and studies on ease of use focused on scoring of written or video patient scenarios. No study assessed ease of use as measured by time to complete assessments or assessor satisfaction.

- The interrater reliability of tGCS and mGCS appears to be high, but evidence was insufficient to determine if there were differences between scales (SOE: Insufficient).

- Three studies found the tGCS associated with a lower proportion of correct scores than the mGCS (differences in proportion of correct scores ranged from 6% to 27%), though the difference was statistically significant in only one study (SOE: Low).
- Three studies found that training or use of a scoring aid increased the proportion of correct scores on both the tGCS and mGCS (increase in proportion of correct scores ranged from 32% to 70%) (SOE: Low).
- Evidence was insufficient to assess effects of patient or assessor characteristics on comparative interrater reliability of the tGCS versus the mGCS (SOE: Insufficient).
- The proportion of correct GCS scores was generally lowest for assessment of patient scenarios with moderate injury severity in three studies, including one study that evaluated the tGCS and the mGCS (SOE: Low).
- Evidence was insufficient to determine effects of level of training or professional background on the proportion of correct scores on the tGCS versus the mGCS (SOE: Insufficient).
- No study evaluated how comparative interrater reliability or ease of use of the tGCS versus the mGCS vary according to assessment setting (SOE: Insufficient).
- One study found agreement between out-of-hospital and ED scores was similar for the tGCS and the mGCS (SOE: Low).

Discussion

Key Findings and Strength of Evidence

Based on head-to-head studies, we found that the tGCS is associated with slightly better predictive utility than the mGCS, based on the AUROC, a measure of discrimination. The tGCS is better able than the mGCS to discriminate people with trauma who undergo neurosurgical intervention, have severe TBI, or undergo emergency intubation from people who do not experience these outcomes. However, the difference in the AUROC on each of these outcomes was small (<0.05). The tGCS was also better than the mGCS at discriminating trauma patients who died during hospitalization from those who survived hospitalization, but the difference in the AUROC was even smaller (0.01) than for non-mortality outcomes. Findings for the tGCS versus the Simplified Motor Scale (SMS) were similar to findings for the tGCS versus the mGCS for non-mortality outcomes, but the SMS performed slightly worse than the mGCS for mortality. Although studies varied in how they defined neurosurgical interventions, severe brain injury, and emergency intubation, findings were generally similar across definitions for these outcomes. Findings for discrimination were robust in sensitivity and subgroup analyses based on the age group analyzed (children vs. adults or mixed), study year (before 2006 or after 2006), or risk of bias ratings. However, sensitivity and subgroup analyses were limited by small numbers of studies, particularly for non-mortality outcomes.

Evidence on how age, type of trauma, intubation status, intoxication status, receipt of field interventions, timing of GCS assessment, or level of training of people administering the GCS impacted predictive utility was limited. Few studies specifically evaluated children or patients with TBI, though those available reported findings similar to studies that evaluated adults or mixed populations of adults and children or mixed trauma patients.

Evidence on interrater reliability and ease of use was limited. Only one study, with methodological limitations and imprecise estimates was included for assessment of patients with

trauma that compared interrater reliability of the tGCS, mGCS, and SMS. Studies that addressed ease of use were limited to those that evaluated whether the measures were scored correctly compared with a reference standard (usually expert assessment). Three studies found that the percentage of correct scores was higher for the mGCS than the tGCS, though in only one study was the difference statistically significant. Limited evidence suggests that errors are more frequent when assessing patient scenarios indicating moderate injury severity (tGCS scores of 9-13).[49,51,54] For both scales, use of a scoring aid or training appears to improve the proportion of correct scores. No study evaluated other measures of ease of use, such as time to complete the assessment or assessor satisfaction.

One study found that agreement between field and ED scores was similar for the tGCS and mGCS.[53] Although differences between field and ED scores were noted for both scales, the study also found that blood pressure readings changed. Therefore, some differences between field and ED scores may accurately reflect changing status of the patient due to receipt of out-of-hospital interventions and evolving clinical status, rather than true lack of agreement.

Applicability

Our findings on predictive utility of different GCS scales appear to have broad applicability to field triage in the United States, as they are based on large studies conducted in U.S. trauma settings in mixed populations of adults and children with various types of trauma. We also restricted study inclusion to studies published after 1995, with most studies conducted in the last 5 to 10 years, suggesting high applicability to use in the context of current trauma systems.

Nonetheless, we identified a number of factors that can impact applicability. Despite the broad applicability of the evidence, its applicability to specific patient populations (e.g., specific type of trauma, age, presence and severity of intoxication, presence of medical comorbidities, and presence of other injuries) is less certain. Limited evidence suggests similar results in children versus mixed populations of adults plus children and in patients with TBI versus mixed trauma populations. No study evaluated how predictive utility varied according to the level or training of field training personnel. In fact, no study that used out-of-hospital scores reported the training of the people administering the GCS. Another factor that could impact applicability is that the performance of the tGCS and mGCS may be different when administered soon after injury (in the field) as opposed to later (after field stabilization and destination decisions have been made and patients have arrived in the ED). However, a number of studies on predictive utility were conducted in ED settings, which is more controlled and easier to study than field settings. Evidence on the predictive utility from studies conducted in the ED may be of limited applicability to field settings. However, we found that predictive utility was similar in studies that utilized out-of-hospital versus ED GCS scores. We also found no clear differences in estimates of predictive utility when we restricted analyses to studies conducted in U.S. settings or to more recent (post 2006) studies, which may be more applicable to current U.S. practice.

Research Recommendations

Head-to-head studies that assess one set of patients with the tGCS and another set with the SMS or mGCS are needed to understand effects on clinical outcomes as well as risk of over- or under-triage. For over- and under-triage, studies should utilize standardized, validated measures. For predictive utility, prospective studies that independently assess patients using the tGCS and the mGCS or SMS would be useful for confirming the findings of the currently available retrospective studies. Studies are needed to better understand the predictive utility in important

subpopulations, including children, older patients, patients with specific types of trauma, and patients who have received field interventions prior to assessment. For patients who are intoxicated or intubated, studies that measure how frequently the tGCS reverts to the mGCS due to the inability to assess the other GCS components would be helpful. Studies that evaluate how the predictive utility of the tGCS compares with the mGCS or SMS varies according to the level of training of assessing personnel in the field are also needed. Finally, studies that assess measures of predictive utility other than discrimination (e.g., calibration, adjusted risk estimates, diagnostic accuracy, risk reclassification) would be useful for providing more complete information regarding predictive utility.

Conclusions

The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use. Research is needed to understand how use of the tGCS versus the mGCS or SMS impacts clinical outcomes and risk of over- or under-triage.

References

1. Centers for Disease Control and Prevention. 10 Leading Causes of Death by Age Group, United States - 2014. http://www.cdc.gov/injury/images/lc-charts/leading_causes_of_death_age_group_2014_1050w760h.gif. Accessed on May 5, 2016.
2. Centers for Disease Control and Prevention. 10 Leading Causes of Injury Deaths by Age Group Highlighting Unintentional Injury Deaths, United States - 2014. http://www.cdc.gov/injury/images/lc-charts/leading_causes_of_injury_deaths_unintentional_injury_2014_1040w740h.gif. Accessed on May 5, 2016.
3. Centers for Disease Control and Prevention. National Hospital Ambulatory Medical Care Survey: 2011 Emergency Department Summary Tables. http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf. Accessed on May 5, 2016.
4. Sasser SM, Hunt RC, Faul M, et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. MMWR Recomm Rep. 2012 Jan 13;61(RR-1):1-20. PMID: 22237112.
5. Faul M, Xu L, Wald MM, et al. Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002–2006. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control: 2010. http://www.cdc.gov/traumaticbraininjury/pdf/blue_book.pdf.
6. Centers for Disease Control and Prevention. Rates of TBI-related emergency department visits, hospitalizations, and deaths-United States, 2001–2010. Injury Prevention & Control: Traumatic Brain Injury. 2016.
7. Committee on Trauma, American College of Surgeons. Resources for Optimal Care of the Injured Patient. 2014. https://www.facs.org/quality%20programs/trauma/vrc/resources.
8. Baxt WG, Moody P. The impact of advanced prehospital emergency care on the mortality of severely brain-injured patients. J Trauma. 1987 Apr;27(4):365-9. PMID: 3573084.
9. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. N Engl J Med. 2006 Jan 26;354(4):366-78. doi: 10.1056/NEJMsa052049. PMID: 16436768.
10. Cook CH, Muscarella P, Praba AC, et al. Reducing overtriage without compromising outcomes in trauma patients. Arch Surg. 2001 Jul;136(7):752-6. PMID: 11448384.
11. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet. 1974 Jul 13;2(7872):81-4. PMID: 4136544.
12. Teasdale G, Maas A, Lecky F, et al. The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol. 2014 Aug;13(8):844-54. doi: 10.1016/S1474-4422(14)70120-6. PMID: 25030516.
13. Marshall LF. Evoked potentials: a decade later. Crit Care Med. 1991 Nov;19(11):1337. PMID: 1935149.
14. Bradshaw D. Report to the Surgeon General: Traumatic Brain Injury Task Force. May 15, 2007. 2008. https://www.hsdl.org/?view&did=482727.
15. Singh B, Murad MH, Prokop LJ, et al. Meta-analysis of Glasgow coma scale and simplified motor score in predicting traumatic brain injury outcomes. Brain Inj. 2013;27(3):293-300. doi: 10.3109/02699052.2012.743182. PMID: 23252405.
16. Brown KM, Hirshon JM, Alcorta R, et al. The implementation and evaluation of an evidence-based statewide prehospital pain management protocol developed using the national prehospital evidence-based guideline model process for emergency medical services. Prehosp Emerg Care. 2014;18 Suppl 1:45-51. doi: 10.3109/10903127.2013.831510. PMID: 24134543.

17. Beskind DL, Stolz U, Gross A, et al. A comparison of the prehospital motor component of the Glasgow coma scale (mGCS) to the prehospital total GCS (tGCS) as a prehospital risk adjustment measure for trauma patients. Prehosp Emerg Care. 2014 Jan-Mar;18(1):68-75. doi: 10.3109/10903127.2013.844870. PMID: 24329032.

18. Healey C, Osler TM, Rogers FB, et al. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. J Trauma. 2003 Apr;54(4):671-8; discussion 8-80. doi: 10.1097/01.ta.0000058130.30490.5d. PMID: 12707528.

19. Ross SE, Leipold C, Terregino C, et al. Efficacy of the motor component of the Glasgow Coma Scale in trauma triage. J Trauma. 1998 Jul;45(1):42-4. PMID: 9680010.

20. Gill M, Windemuth R, Steele R, et al. A comparison of the Glasgow Coma Scale score to simplified alternative scores for the prediction of traumatic brain injury outcomes. Ann Emerg Med. 2005 Jan;45(1):37-42. PMID: 15635308.

21. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem. 2008 Jan;54(1):17-23. doi: 10.1373/clinchem.2007.096529. PMID: 18024533.

22. Menegazzi JJ, Davis EA, Sucov AN, et al. Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. J Trauma. 1993 Jan;34(1):46-8. PMID: 8437195.

23. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. Ann Emerg Med. 2004 Feb;43(2):215-23. doi: 10.1016/S019606440300814X. PMID: 14747811.

24. Neville AL, Nemtsev D, Manasrah R, et al. Mortality risk stratification in elderly trauma patients based on initial arterial lactate and base deficit levels. Am Surg. 2011;77(10):1337-41 5p. PMID: 104607567.

25. National Emergency Medical Services Workforce Data Definitions. (Prepared by the American Institutes for Research through contract DTNH22-080F-00122.) National Highway Traffic Safety Administration. Washington, DC: February 2013. www.nhtsa.gov/staticfiles/nti/pdf/811720.pdf.

26. Badjatia N, Carney N, Crocco TJ, et al. Guidelines for prehospital management of traumatic brain injury 2nd edition. Prehosp Emerg Care. 2008;12 Suppl 1:S1-52. doi: 10.1080/10903120701732052. PMID: 18203044.

27. Bazzoli GJ, Madura KJ, Cooper GF, et al. Progress in the development of trauma systems in the United States. Results of a national survey. JAMA. 1995 Feb 1;273(5):395-401. PMID: 7823385.

28. Acker SN, Ross JT, Partrick DA, et al. Glasgow motor scale alone is equivalent to Glasgow Coma Scale at identifying children at risk for serious traumatic brain injury. J Trauma Acute Care Surg. 2014 Aug;77(2):304-9. doi: http://dx.doi.org/10.1097/TA.0000000000000300. PMID: 25058258.

29. Al-Salamah MA, McDowell I, Stiell IG, et al. Initial emergency department trauma scores from the OPALS study: the case for the motor score in blunt trauma. Acad Emerg Med. 2004 Aug;11(8):834-42. PMID: 15289188.

30. Brown JB, Forsythe RM, Stassen NA, et al. Evidence-based improvement of the National Trauma Triage Protocol: The Glasgow Coma Scale versus Glasgow Coma Scale motor subscale. J Trauma Acute Care Surg. 2014 Jul;77(1):95-102; discussion 1-2. doi: http://dx.doi.org/10.1097/TA.0000000000000280. PMID: 24977762.

31. Caterino JM, Raubenolt A. The prehospital simplified motor score is as accurate as the prehospital Glasgow coma scale: analysis of a statewide trauma registry. Emerg Med J. 2012 Jun;29(6):492-6. doi: http://dx.doi.org/10.1136/emj.2010.110437. PMID: 21795294.

32. Caterino JM, Raubenolt A, Cudnik MT. Modification of Glasgow Coma Scale criteria for injured elders. Acad Emerg Med. 2011 Oct;18(10):1014-21. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01164.x. PMID: 21951715.

33. Cicero MX, Cross KP. Predictive value of initial Glasgow coma scale score in pediatric trauma patients. Pediatr Emerg Care. 2013 Jan;29(1):43-8. doi: http://dx.doi.org/10.1097/PEC.0b013e31827b52bf. PMID: 23283262.

34. Corrigan JD, Kreider S, Cuthbert J, et al. Components of traumatic brain injury severity indices. J Neurotrauma. 2014 Jun;31(11):1000-7. doi: http://dx.doi.org/10.1089/neu.2013.3145. PMID: 24521197.

35. Davis DP, Serrano JA, Vilke GM, et al. The predictive value of field versus arrival Glasgow Coma Scale score and TRISS calculations in moderate-to-severe traumatic brain injury. J Trauma. 2006 May;60(5):985-90. PMID: 16688059.

36. Eken C, Kartal M, Bacanli A, et al. Comparison of the Full Outline of Unresponsiveness Score Coma Scale and the Glasgow Coma Scale in an emergency setting population. Eur J Emerg Med. 2009 Feb;16(1):29-36. doi: http://dx.doi.org/10.1097/MEJ.0b013e32830346ab. PMID: 19106717.

37. Gill M, Steele R, Windemuth R, et al. A comparison of five simplified scales to the out-of-hospital Glasgow Coma Scale for the prediction of traumatic brain injury outcomes. Acad Emerg Med. 2006 Sep;13(9):968-73. PMID: 16894005.

38. Haukoos JS, Gill MR, Rabon RE, et al. Validation of the Simplified Motor Score for the prediction of brain injury outcomes after trauma. Ann Emerg Med. 2007 Jul;50(1):18-24. PMID: 17113193.

39. Holmes JF, Palchak MJ, MacFarlane T, et al. Performance of the pediatric glasgow coma scale in children with blunt head trauma. Acad Emerg Med. 2005 Sep;12(9):814-9. PMID: 16141014.

40. Johnson DL, Krishnamurthy S. Send severely head-injured children to a pediatric trauma center. Pediatr Neurosurg. 1996 Dec;25(6):309-14. PMID: 9348151.

41. Leijdesdorff HA, van Dijck JT, Krijnen P, et al. Injury pattern, hospital triage, and mortality of 1250 patients with severe traumatic brain injury caused by road traffic accidents. J Neurotrauma. 2014 Mar 1;31(5):459-65. doi: http://dx.doi.org/10.1089/neu.2013.3111. PMID: 24093437.

42. Majdan M, Steyerberg EW, Nieboer D, et al. Glasgow coma scale motor score and pupillary reaction to predict six-month mortality in patients with traumatic brain injury: comparison of field and admission assessment. J Neurotrauma. 2015 Jan 15;32(2):101-8. doi: http://dx.doi.org/10.1089/neu.2014.3438. PMID: 25227136.

43. Nesiama JA, Pirallo RG, Lerner EB, et al. Does a prehospital Glasgow Coma Scale score predict pediatric outcomes? Pediatr Emerg Care. 2012 Oct;28(10):1027-32. doi: http://dx.doi.org/10.1097/PEC.0b013e31826cac31. PMID: 23023472.

44. Reisner A, Chen X, Kumar K, et al. Prehospital heart rate and blood pressure increase the positive predictive value of the Glasgow Coma Scale for high-mortality traumatic brain injury. J Neurotrauma. 2014 May 15;31(10):906-13. doi: http://dx.doi.org/10.1089/neu.2013.3128. PMID: 24372334.

45. Thompson DO, Hurtado TR, Liao MM, et al. Validation of the Simplified Motor Score in the out-of-hospital setting for the prediction of outcomes after traumatic brain injury. Ann Emerg Med. 2011 Nov;58(5):417-25. doi: 10.1016/j.annemergmed.2011.05.033. PMID: 21803448.

46. Timmons SD, Bee T, Webb S, et al. Using the abbreviated injury severity and Glasgow Coma Scale scores to predict 2-week mortality after traumatic brain injury. J Trauma. 2011 Nov;71(5):1172-8. doi: http://dx.doi.org/10.1097/TA.0b013e31822b0f4b. PMID: 22071922.

47. Van de Voorde P, Sabbe M, Rizopoulos D, et al. Assessing the level of consciousness in children: a plea for the Glasgow Coma Motor subscore. Resuscitation. 2008 Feb;76(2):175-9. doi: 10.1016/j.resuscitation.2007.07.007. PMID: 17728046.

48. Arbabi S, Jurkovich GJ, Wahl WL, et al. A comparison of prehospital and hospital data in trauma patients. J Trauma. 2004 May;56(5):1029-32. PMID: 15179242.

49. Bledsoe BE, Casey MJ, Feldman J, et al. Glasgow Coma Scale Scoring is Often Inaccurate. Prehosp Disaster Med. 2015 Feb;30(1):46-53. doi: http://dx.doi.org/10.1017/S1049023X14001289. PMID: 25489727.

50. Dinh MM, Oliver M, Bein K, et al. Level of agreement between prehospital and emergency department vital signs in trauma patients. Emerg Med Australas. 2013 Oct;25(5):457-63. doi: http://dx.doi.org/10.1111/1742-6723.12126. PMID: 24099376.

51. Feldman A, Hart KW, Lindsell CJ, et al. Randomized controlled trial of a scoring aid to improve glasgow coma scale scoring by emergency medical services providers. Ann Emerg Med. 2015;65(3):325-9.e2. PMID: 25199613.

52. Heim C, Schoettker P, Gilliard N, et al. Knowledge of Glasgow coma scale by air-rescue physicians. Scand J Trauma Resusc Emerg Med. 2009;17:39. doi: 10.1186/1757-7241-17-39. PMID: 19723331.

53. Kerby JD, MacLennan PA, Burton JN, et al. Agreement between prehospital and emergency department glasgow coma scores. J Trauma. 2007 Nov;63(5):1026-31. PMID: 17993947.

54. Lane PL, Baez AA, Brabson T, et al. Effectiveness of a Glasgow Coma Scale instructional video for EMS providers. Prehosp Disaster Med. 2002 Jul-Sep;17(3):142-6. PMID: 12627917.

55. Takahashi C, Okudera H, Origasa H, et al. A simple and useful coma scale for patients with neurologic emergencies: the Emergency Coma Scale. Am J Emerg Med. 2011 Feb;29(2):196-202. doi: http://dx.doi.org/10.1016/j.ajem.2009.09.018. PMID: 20825789.

Introduction

Background

Nature and Burden of Trauma

Unintentional injuries are the leading cause of death among people in the United States ages 1 to 44, and the third leading cause among people ages 45 to 64.[1] Among all age groups, motor vehicle crashes are the first or second leading cause of unintentional injury death.[2] In 2011, there were approximately 40,000,000 emergency department (ED) visits for injuries; of these approximately 2.5 million were due to trauma complications and unspecified injuries.[3] Approximately 18 percent of patients seen in the ED for an injury were transported by Emergency Medical Services (EMS) personnel.[4] Traumatic brain injury (TBI) is an important subset of trauma. Among an estimated 1.7 million annual cases of TBI, there are 52,000 deaths and 275,000 hospitalizations.[5] TBI is a contributing factor to about one third of injury related deaths in the United States. From 2001 to 2010, the rate of TBI-related ED visits increased from 421 to 716 per 100,000,[6] though the rate of deaths declined from 18.5 to 17.1 per 100,000 people.

Field Triage of Patients with Trauma

Field triage by EMS is a critical aspect of trauma systems, as it helps to identify seriously injured patients for transport to major trauma centers,[7] which have been shown to improve survival in such patients.[8,9] Appropriate decisions regarding transport are crucial because management of severely injured patients in a Level I or a Level II trauma center has been shown to be associated with improved clinical outcomes.[9] On the other hand, unnecessarily triaging patients to high level trauma care who do not require it may represent an inefficient use of staff and resources.[10]

EMS providers must rapidly triage individuals who have undergone trauma in challenging environments. Therefore, EMS providers must have assessment tools that are easy to use, reliable, and accurate. A key component of field triage for patients with suspected serious injury is level of consciousness assessment.[4] The Glasgow Coma Scale (GCS)[11,12] is an instrument widely used for assessment of consciousness at the site of injury, in emergency departments, and in hospitals, as well as to monitor progress or deterioration during treatment.[13] The GCS consists of three items (components): eye (scored 1 to 4), verbal (scored 1 to 5), and motor (scored 1 to 6). Scores on each of these components are added to obtain the total Glasgow Coma Scale (tGCS) score, ranging from 3 to 15. Lower scores on the tGCS indicate lower levels of consciousness, generally correlating with more severe injury associated with poorer prognosis and requiring more intensive care. For patients with TBI, scores of 3 to 8 are generally considered to denote severe head injury, 9 to 12 moderate, and 13 to 15 mild.[14] The 2011 field triage guidelines from the Centers for Disease Control and Prevention (CDC) National Expert Panel recommend transporting patients with tGCS scores of 13 or less to facilities providing the highest level of trauma care.[4]

In some circumstances (e.g., trauma victims who are intoxicated, intubated, or whose other injuries influence response) it may not be possible to accurately assess the verbal and eye components of the tGCS. In these cases, assessments may be primarily based on the motor component of the Glasgow Coma Scale (mGCS) alone.[11,15-17] In addition, the mGCS has been

proposed for assessment of trauma patients even when the tGCS can be obtained, since only one item is assessed, potentially increasing ease of use in the field.[18] mGCS scores of 5 or less are considered an indication of patients with severe injury.[18,19] The Simplified Motor Score (SMS) has been proposed as a streamlined alternative to the mGCS; it is assessed on a three point scale (scored 0 to 2, with a score of 0 corresponding to 1 to 4 on the mGCS, 1 corresponding to 5 on the mGCS, and 2 corresponding to 6 on the mGCS).[20]

Decisions regarding the use of field triage instruments for level of consciousness should be based on how they perform in comparison to the tGCS. The ultimate goal of selecting one risk prediction instrument over another is to improve clinical outcomes (e.g., mortality). However, information on clinical outcomes is often lacking and decisions about their use must often be based on how they perform on intermediate outcomes. Intermediate outcomes include measures of over- or under-triage (i.e., the degree to which patients are unnecessarily transported to a Level I or II trauma center [over-triage] or not transported to a Level I or II trauma center [under-triage]) or predictive utility, as assessed using measures of discrimination (ability of an instrument to distinguish people with the disease from those without), calibration (how well predicted risk correlates with actual risk), standard measures of diagnostic accuracy (e.g., sensitivity, specificity, and predictive values), or adjusted risk estimates (e.g., odds ratio, relative risk, hazards ratio).[21] Other factors that could inform selection of field triage risk assessment instruments include intra- and interrater reliability and ease of use (e.g., time to administer the instrument and amount of missing data).[12,22,23]

A number of factors could impact the performance of field assessment instruments. These include variability in patient populations (e.g., type of trauma, demographic characteristics, presence and severity of intoxication, and medical comorbidities), level of training and certification of administering personnel (e.g., Emergency Medical Technician [EMT],[24] EMT-Intermediate, Advanced EMT/Paramedic, physician, or nurse[25]), receipt of field interventions (e.g., medications, intubation), setting (e.g., country, urban vs. rural) or timing of assessment relative to injury occurrence. Evidence about field triage instruments frequently relies on extrapolation from studies conducted in EDs, as this environment is more controlled and easier to study.[26] However, the performance of the tGCS and mGCS may be different when administered soon after injury by EMS personnel in the field as opposed to later by ED personnel, after destination decisions have already been made and patients have been stabilized with initial interventions.

During the development of field triage guidelines and algorithms by the CDC National Expert Panel in 2011,[4] use of the mGCS was considered a way to potentially simplify field triage. However, the mGCS was not adopted due in part to lack of evidence about the comparative accuracy and reliability of the mGCS relative to the tGCS. However, more evidence is now available on the mGCS.

Rationale for Review

The purpose of this report is to systematically review the currently available evidence on the comparative predictive utility, reliability, and ease of use of the tGCS and mGCS in field assessment of trauma (with or without TBI), as well as comparative effects on clinical outcomes and early critical resource use. This review provides a synthesis of currently available evidence and gaps in evidence that may be helpful to inform clinical practice and guideline development for field triage of trauma by EMS personnel. The review is the first step of a larger Federal effort

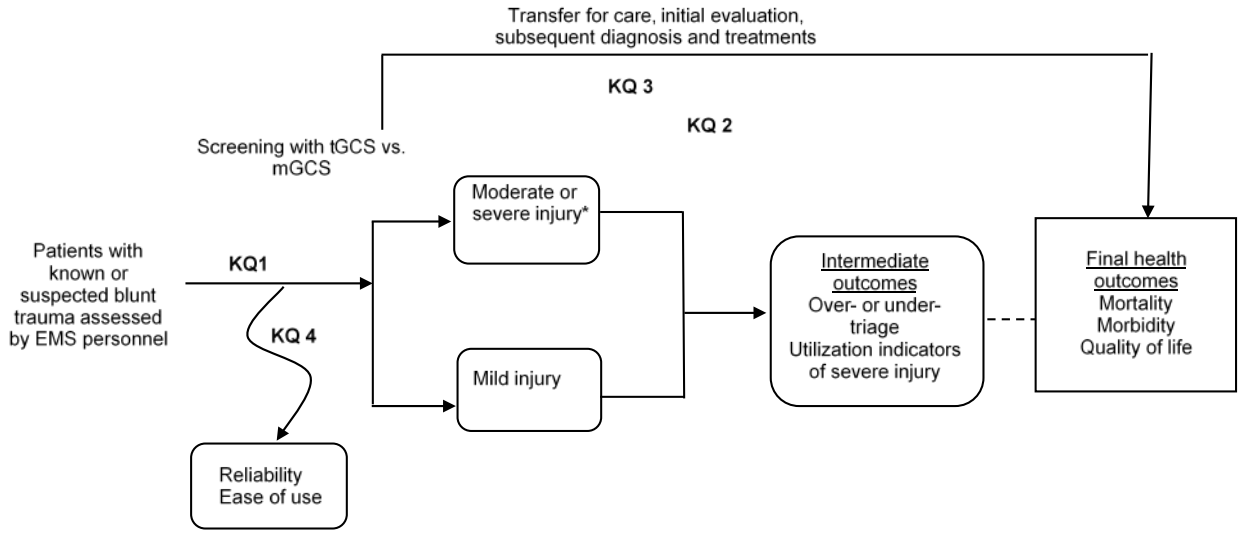
to systematically examine the evidence base about out-of-hospital triage decisionmaking and transport of trauma patients, and inform future updates to the Field Triage Guidelines.[4]

Scope of Review and Key Questions

The research questions were initially developed by the National Highway Traffic Safety Administration and revised with input from a Technical Expert Panel. The Key Questions focus on predictive utility, over- and under-triage, clinical outcomes of the tGCS versus the mGCS or SMS, as well as reliability and ease of use. Key Question 1 addresses the predictive utility of the tGCS compared with the mGCS for predicting clinical outcomes (mortality, morbidity). In addition, Key Question 1 addresses the predictive utility of the tGCS versus the mGCS on markers of injury severity, as indicated by the injury severity score and utilization markers for severe injury (receipt of neurosurgical interventions such as early surgery or intracranial pressure monitoring) and as a marker of need for tertiary trauma care. Key Question 1 does not directly assess the utility of the tGCS compared with the mGCS for predicting the likelihood that a patient receives tertiary trauma care, since the triage assessment is one of the factors used to determine who is transported to tertiary trauma care; therefore, receipt of tertiary trauma care does not represent a marker of injury severity independent from the GCS score. Key Question 2 addresses the impact of the tGCS compared with the mGCS on rates of over- and under-triage, as measured by initial EMS transport to a lower or higher level of care in combination with injury severity or critical resource use, an intermediate outcome. Measuring over- and under-triage is a challenge because factors other than findings on field triage assessment scales, including other patient characteristics (e.g., mechanism of injury, hemodynamic instability, respiratory distress, comorbidities), geographic proximity, and availability of resources, also impact triage decisions.[10] Some over-triage may be acceptable in order to prevent under-triage,[7,27] which may be more likely to result in adverse clinical outcomes,[28,29] while over-triage may primarily represent inefficient use of resources and increased costs[30,31] without necessarily adversely impacting clinical outcomes.[32] Therefore, results for Key Question 2 must be interpreted with caution. Key Question 3 addresses the impact of the tGCS compared with the mCGS on clinical outcomes. Key Question 4 addresses the reliability and ease of use of the tGCS compared with the mGCS. For each Key Question, a subquestion addresses potential modifiers of treatment effect, including patient age or other patient characteristics, the training and background of the person administering the instrument, and the timing/setting of assessment. The analytic framework (Figure 1) and Key Questions used to guide this review are shown below.

The analytic framework shows the target populations, interventions, and health outcomes examined, with numbers corresponding to the Key Questions.

Figure 1. Analytic framework



EMS=emergency medical services; KQ=key question; mGCS=motor Glasgow Coma Scale; tGCS=total Glasgow Coma Scale; vs.=versus

*Based on tGCS score of ≤13 or mGCS score of ≤5

The report focuses on the following Key Questions:

Key Question 1. In patients with known or suspected trauma, what is the predictive utility of the tGCS compared with the mGCS for predicting mortality, morbidity, Injury Severity Score of 16 or greater, head Abbreviated Injury Scale score greater than 2 or greater than 3, presence of intracranial hemorrhage, and utilization indicators of severe injury (e.g., receipt of intracranial monitoring within 48 hours of admission, receipt of surgery within 12 hours of admission, or early intubation [in the field or immediately upon presentation to the ED])?

Key Question 1a. How does predictive utility vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medications in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 2. In patients with known or suspected trauma, what are the comparative effects of the tGCS compared with the mGCS on over- and under-triage (e.g., proportion of patients misclassified with regard to measures of injury severity or need for early interventions and transport to a lower vs. higher level of care)?

Key Question 2a. How do effects on clinical decisionmaking vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 3. In patients with known or suspected trauma, what is the comparative effectiveness of the tGCS compared with the mGCS on clinical outcomes (e.g., mortality, morbidity, quality of life)?

Key Question 3a. How do effects on clinical outcomes vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Question 4. In patients with known or suspected trauma, what is the comparative reliability (e.g., interrater and intra-rater kappa) and ease of use (e.g., time to complete, amount of missing data, user reported satisfaction) of the tGCS compared with the mGCS score?

Key Question 4a. How do comparative reliability and ease of use vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Methods

The methods for this Comparative Effectiveness Review (CER) follow the guidance in the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews.[33]

Scope Development

The initial Key Questions were provided by the National Highway Traffic Safety Administration (NHTSA). The Key Questions were further developed and the final protocol developed with additional input from NHTSA and a Technical Expert Panel (TEP) convened for this report. The TEP consisted of eight experts in adult and pediatric field triage, Emergency Medical Services, trauma surgery, general surgery, critical care and emergency medicine. TEP members disclosed financial and other conflicts of interest prior to participation. The AHRQ Task Order Officer and the investigators reviewed the disclosures and determined that the TEP members had no conflicts of interest that precluded participation.

The final protocol was posted on the AHRQ Web site on February 28, 2016 at: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2187. The protocol was also registered in the PROSPERO international database of prospectively registered systematic reviews (registration number CRD42016035944).[34]

Literature Search Strategy

A research librarian conducted searches in the Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, CINAHL (Cumulative Index to Nursing and Allied Health Literature), PsycINFO, HAPI (Health & Psychosocial Instruments), and Ovid MEDLINE (January, 1995 to February, 2016), limiting to English-language abstracts. Search strategies are provided in Appendix A. We restricted search start dates to January, 1995 to improve applicability to current U.S. trauma care; only five states had fully implemented trauma systems in the early 1990s.[35] In addition, the first studies to compare the predictive utility of the motor component of the Glasgow Coma Scale (mGCS) versus the total score of the Glasgow Coma Scale (tGCS) were published in 1998 and 2003.[18,19]

We also hand-searched the reference lists of relevant studies and searched for unpublished studies in ClinicalTrials.gov.

Study Selection

We developed criteria for inclusion and exclusion of studies based on the Key Questions and the populations, interventions, comparators, outcomes, timing, types of studies, and setting (PICOTS) approach, in accordance with the AHRQ Methods Guide.[33] Inclusion and exclusion criteria are summarized below and described in more detail in Appendix B. Abstracts were reviewed by two investigators, and all citations deemed potentially appropriate for inclusion by at least one of the reviewers was retrieved for full-text review. Two investigators then independently reviewed all full-text articles for final inclusion. Inclusion was restricted to English-language articles. Discrepancies were resolved by discussion and consensus, with a third investigator if necessary.

A list of the included studies appears in Appendix C; a list of excluded studies and primary reasons for exclusion can be found in Appendix D.

Population and Conditions of Interest

For all Key Questions we included studies of children and adults with known or suspected trauma. Although the population of interest was patients with blunt trauma, we included studies of general trauma patients. We excluded studies of individuals without trauma, studies that focused on non-blunt trauma patients, or studies of patients with and without trauma in which the proportion without trauma was less than 10 percent and results were not reported separately for patients with trauma.

Interventions, Comparisons, and Study Designs

We focused on studies of the tGCS, the mGCS, and the Simplified Motor Score (SMS). For studies evaluating measures of diagnostic accuracy (sensitivity, specificity, predictive values), we focused on studies that used standard cutoff scores of 13 or less on the tGCS, 5 or less on the mGCS, or 1 or 0 on the SMS to indicate people who require high level trauma care,[4,36] but also included studies that used alternative cutoffs or modifications of the tGCS and mGCS. We excluded studies that evaluated the utility of mGCS or tGCS in combination with other predictors in a multi-item risk assessment or triage instrument.

For all Key Questions, we included cohort studies and randomized trials that directly compared the tGCS versus the mGCS or SMS. For Key Question 4 (reliability and ease of use), we also included cross-sectional studies and studies that assessed one of these scales. For Key Question 1a (predictive utility) we included studies that assessed one of these scales if they addressed one of the subpopulations specified in the Key Questions (e.g., children, intoxicated, intubated, traumatic brain injury [TBI] patients) not addressed well in the head-to-head studies. We restricted Key Questions 2 and 3 to head-to-head studies because it is not possible to isolate the effects of risk assessment scales on over- or under-triage or clinical outcomes from single-arm studies, given the large number of other factors that impact these outcomes.

Outcomes

For Key Question 1, we included measures of predictive utility for mortality, morbidity, markers of severe injury (e.g., Abbreviated Injury Scale [AIS] score of ≥4 or Injury Severity Score [ISS] of ≥16[37]) or utilization indicators of severe injury[38] (e.g., receipt of intracranial monitoring within 48 hours of admission, receipt of surgery within 12 hours of admission, or receipt of early intubation [in the field or immediately upon arrival to the emergency department {ED}]), as measured by diagnostic accuracy, adjusted risk estimates, measures of discrimination (e.g., the c-index), measures of calibration (e.g., the Hosmer-Lemeshow test), or risk reclassification rates.[39]

For Key Question 2, we included studies that reported the proportion of patients who were over- or under-triaged (e.g., the proportion transferred to a higher or lower level of care).[39]

For Key Question 3, we included studies that reported clinical outcomes, including mortality (prior to hospital arrival, in the ED, or after hospital admission), measures of morbidity, including cognitive impairment and medical complications related to trauma, and quality of life, including functional capacity at discharge or followup.

For Key Question 4, we included outcomes that assessed reliability (e.g., interrater and intra-rater kappa) or ease of use (e.g., time to complete, measures of missing data, user reported satisfaction).

Timing and Setting

For all Key Questions we included prospective and retrospective studies in which the tGCS, mGCS, or SMS were administered soon after injury (conducted in the field/out-of-hospital setting by Emergency Medical Services personnel) or immediately upon arrival to the ED, or that were based on trauma registry data collected in the field or in the ED. We excluded studies in which the Glasgow Coma Scale (GCS) was administered after more than 4 hours in the ED or hospital or when it was administered after hospital admission. We also excluded studies conducted in the developing world, which may have limited applicability to U.S. trauma care settings.

Data Abstraction and Data Management

A single investigator abstracted information on study design, year, geographic location, patient characteristics (i.e. demographics, type and mechanism of trauma, type of injury, tGCS scores, severity of injury, intoxication status, systolic blood pressure, intubation or receipt of medication in the field), the proportion of patients who experienced outcomes, which triage instrument was used, timing of triage assessment, cutoff scores used (for studies that evaluated sensitivity and specificity), the training and experience of the person administering the GCS, assessment setting (in the field or upon ED presentation), and results relevant to each Key Question. All data abstractions were reviewed by a second investigator for accuracy and discrepancies were resolved through discussion and consensus.

Assessment of Methodological Risk of Bias of Individual Studies

We assessed risk of bias of included studies using predefined criteria. Our methods for assessing risk of bias are based on the recommendations in the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews.[33] For Key Question 1 (predictive utility), we applied the Quality in Prognostic Studies (QUIPS) tool for prognostic studies.[40] The QUIPS tool includes domains on study participation, study attrition, prognostic factor measurement, outcomes measurement, study confounding, and statistical analysis and reporting. For Key Question 4 (reliability and ease of use), we assessed risk of bias using criteria adapted from the Quality Assessment of Diagnostic Accuracy Studies (QUADAS).[41] This includes criteria about patient selection, whether raters were blinded to other ratings, how the scores from different assessments were compared and the situation and timing of measurement. Two investigators independently assessed risk of bias for each study. Differences were resolved by discussion; we used a third rater to resolve discrepancies if needed (Appendix E). No study met inclusion criteria for Key Questions 2 or 3.

Studies rated “low” risk of bias have the least risk of bias, and their results are generally considered more valid than studies with the same study design but more flaws. For example, low risk of bias studies on predictive utility select all or a random subset of patients who meet pre-defined criteria, report low attrition, perform the risk assessment scale in all patients, measure

outcomes accurately and in all patients, assess and measure important confounders, and use appropriate statistical methods and avoid selective reporting of results.

Studies rated "moderate" risk of bias are susceptible to some bias, though not enough to necessarily invalidate the results. These studies may not meet all the criteria for "low" risk of bias rating, but do not have flaws likely to cause major bias. The study may also be missing information, making it difficult to assess limitations and potential problems. The "moderate" risk of bias category is broad, and studies with this rating will vary in their strengths and weaknesses. The results of some "moderate" risk of bias studies are likely to be valid, while others may be only possibly valid.

Studies rated "high" risk of bias have significant flaws that may invalidate the results. They may have a serious or "fatal" flaw or set of flaws in design, analysis, or reporting; large amounts of missing information; or discrepancies in reporting. The results of these studies are at least as likely to reflect flaws in the study design as the true difference between the compared interventions. We did not exclude studies rated high risk of bias a priori, but performed sensitivity analyses in which such studies were excluded.

Assessing Research Applicability

Applicability is defined as the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under "real-world" conditions.[33] It is an indicator of the extent to which research included in a review might be useful for informing clinical decisions in specific situations. Because applicability depends on the perspective of the user of the review, we did not assign a rating for applicability (such as "high" or "low"). Rather, factors important for understanding the applicability of studies were recorded, such as population characteristics (age, type of trauma, intoxication status), setting (U.S. vs. other country, out-of-hospital vs. ED assessment), and type and level of training of people administering the GCS were recorded and assessed in subgroup and sensitivity analyses.[42] We also recorded the funding source for studies. Most studies on predictive utility reported the area under the receiver operating characteristic (AUROC), a measure of discrimination.[43,44] The AUROC value represents the probability that a patient who experiences an outcome will have a higher score on the triage scale than a person who does not experience the outcome. We did not identify published recommendations on how to interpret the magnitude of differences in the AUROC value. Therefore, we defined a small difference in the AUROC *a priori* as a difference of less than 0.05, moderate as a difference of 0.05 to 0.10, and large as a difference of greater than 0.10.

Data Synthesis and Rating the Body of Evidence

We performed random effects meta-analysis to calculate pooled differences on the AUROC from head-to-head studies[45] of the tGCS versus the mGCS or SMS using the DerSimonian-Laird model with Stata/IC 13.1 (StataCorp LP, College Station, TX). We measured statistical heterogeneity using the I^2 statistic. The DerSimonian-Laird estimator can result in confidence intervals (CIs) for the pooled estimate that are too narrow, particularly when statistical heterogeneity is present.[46] Therefore, we repeated analyses using an alternative random effects model, the profile likelihood method, which may provide more accurate confidence limits. Most studies reported estimated AUROCs with associated 95 percent CIs. When a study only reported the point estimate of AUROC without providing a 95 percent CI or a standard error, we imputed the standard error using the average standard error from other studies in the same meta-analysis.

In addition, in all studies, the mGCS or SMS scores were derived from the tGCS and applied to the same patient population (i.e. the risk assessment scales were not applied independently). To account for this non-independence, we assumed a correlation of 0.5 when comparing the tGCS versus the mGCS, or SMS in the primary analysis. Two studies[47,48] reported data that allowed us to calculate the correlations between the AUROC for the tGCS and the mGCS or SMS, which ranged from 0.5 to 0.9 depending on the outcome and comparison. Therefore, 0.5 is a conservative assumption for the correlation. A high correlation is expected given mGCS or SMS scores are a subset of tGCS applied on the same population. Sensitivity analyses were conducted assuming correlations of 0.3 and 0.8; results were similar and not separately reported.

Primary analyses were stratified by the age group evaluated in the study (children vs. adults or mixed populations). We performed additional sensitivity and subgroup analyses based on timing of GCS assessment (field vs. ED), study dates (all data collected after 2006 or some or all data collected prior to 2006), country (United States vs. other) and risk of bias rating. For the primary analysis, we included studies conducted using the National Trauma Data Bank (NTDB). In 2012, 805 hospitals submitted data to the NTDB.[49] Because populations evaluated in single trauma centers or systems could be included (in part or in full) in the NTDB, we performed a sensitivity analysis in which NTDB studies were excluded. For the primary analysis, we included multiple studies from the same trauma center or system unless there was clearly complete overlap in the populations assessed. In sensitivity analyses, we restricted analyses to studies from each trauma center that utilized field GCS scores; if multiple studies utilized field GCS scores we utilized the study that evaluated more recent data.

Grading the Body of Evidence for Each Key Question

For all comparisons and outcomes we assessed the strength of evidence using the approach described in the AHRQ Methods Guide (Appendix F),[33,50] based on the overall risk of bias (graded low, moderate, or high); the consistency of results across studies (graded consistent, inconsistent, or unable to determine when only one study was available); the directness of the evidence linking the intervention and health outcomes (graded direct or indirect); the precision of the estimate of effect, based on the number and size of studies and CIs for the estimates (graded precise or imprecise); and reporting bias (suspected or undetected). Assessments of reporting bias were based on whether studies defined and reported primary outcomes and whether we identified relevant unpublished studies.

Based on our assessments on the domains described above, we graded the strength of evidence for each Key Question using the four key categories recommended in the AHRQ Methods Guide.[51] Randomized controlled trials and cohort studies on predictive utility started as "high" strength of evidence and graded down based on the presence of deficiencies in the domains. Because observational studies on predictive utility started as high, we did not consider factors for upgrading such as dose-response relationship, magnitude of effects, or impact of plausible confounders. A "high" grade indicates high confidence that the evidence reflects the true effect and that further research is very unlikely to change our confidence in the estimate of effect. A "moderate" grade indicates moderate confidence that the evidence reflects the true effect and further research may change our confidence in the estimate of effect and may change the estimate. A "low" grade indicates low confidence that the evidence reflects the true effect and further research is likely to change the confidence in the estimate of effect and is likely to change the estimate. An "insufficient" grade indicates evidence either is unavailable or is too

limited to permit any conclusion, due to the availability of only high risk of bias studies, extreme inconsistency, or extreme imprecision.

See Appendix G for the strength of evidence table.

External Review

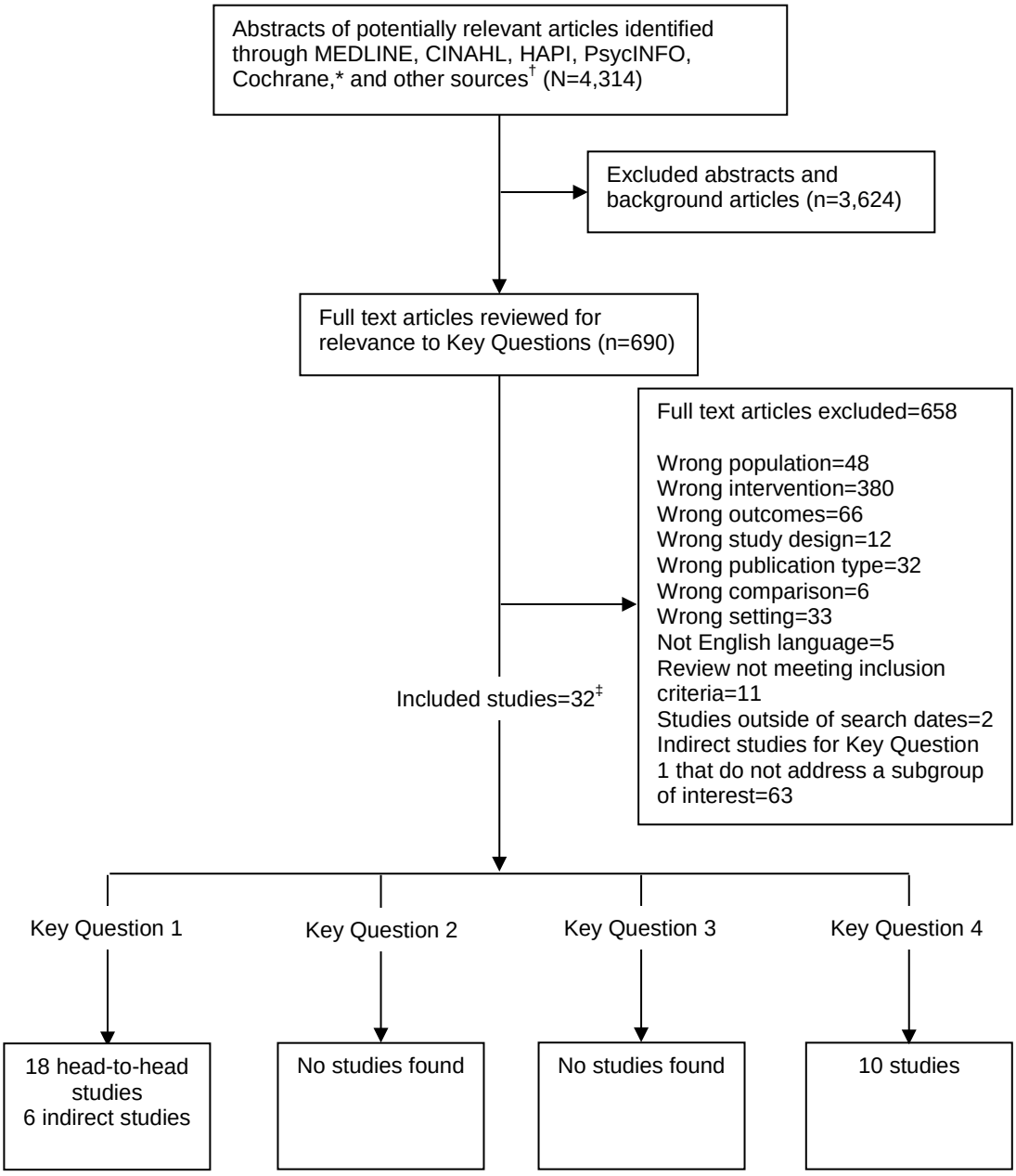
Peer reviewers with expertise in trauma and triage have been invited to provide written comments on the draft report. The AHRQ Task Order Officer and an Evidence-based Practice Center Associate Editor will also provide comments and editorial review. The draft report will be posted on the AHRQ Web site for 4 weeks for public comment. A disposition of comments report with authors' responses to the peer and public review comments will be posted after publication of the final CER on the public Web site.

Results

Results of Literature Search

Results of the literature search and selection process are summarized in the literature flow diagram (Figure 2). Database searches resulted in 4,306 potentially relevant citations. After dual review of abstracts and titles, 690 articles were selected for full-text review. After dual review of full text articles, 32 studies were included. Data extraction and risk of bias assessment tables for included studies by Key Question are available in Appendixes H through L.

Figure 2. Literature flow diagram



CINAHL=The Cumulative Index to Nursing and Allied Health Literature; HAPI=Health and Psychosocial Instruments; MEDLINE=Medical Literature Analysis and Retrieval System Online; n=sample size

*Cochrane databases include the Cochrane Central Register of Controlled Trials and the Cochrane Database of Systematic Reviews.
†Other sources include reference list, experts, etc.
‡Two studies were used for Key Question 1 and Key Question 4

Key Question 1. In patients with known or suspected trauma, what is the predictive utility of the total Glasgow Coma Scale (tGCS) compared with the motor GCS (mGCS) score for predicting mortality, morbidity, Injury Severity Score (ISS) of 16 or greater, head Abbreviated Injury Scale (AIS) score greater than 2 or greater than 3, presence of intracranial hemorrhage, and utilization indicators of severe injury (e.g., receipt of intracranial monitoring within 48 hours of admission, receipt of surgery within 12 hours of admission, or early intubation [in the field or immediately upon presentation to the emergency department {ED}])?

Key Points

- In-hospital mortality
  - For the tGCS versus the mGCS, the pooled area under the receiver operating characteristic (AUROC) was 0.886 (95% confidence interval [CI] 0.863 to 0.908) versus 0.864 (95% CI 0.839 to 0.890), respectively with a pooled mean difference of 0.013 (95% CI 0.007 to 0.019; I^2=59%), based on 11 studies (strength of evidence [SOE]: Moderate).
  - For the tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5), differences in sensitivity ranged from 0 percent to 3 percent; difference in specificity ranged from 0 percent to 5 percent in favor of the mGCS, though the CIs overlapped in each study (SOE: Low).
  - For the tGCS versus the Simplified Motor Score (SMS), the pooled AUROC was 0.884 (95% CI 0.852 to 0.916) versus 0.840 (95% CI 0.802 to 0.878), respectively, for a mean difference of 0.030 (95% CI 0.024 to 0.036, I^2=0%), based on five studies (SOE: Moderate).
  - One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity versus the SMS (cutoff of ≤1) (75%, 95% CI 73 to 76 vs. 72%, 95% CI 70 to 74) and slightly lower specificity (88%, 95% CI 87 to 88 vs. 89%, 95% CI 89 to 87) (SOE: Low).
  - For the mGCS versus the SMS, the mean difference in the pooled AUROC was 0.014 (95% CI 0.006 to 0.021, I2=0%), based on four studies (SOE: Moderate).
- Neurosurgical intervention
  - For the tGCS versus the mGCS, the pooled AUROC was 0.798 (95% CI 0.754 to 0.842) versus 0.769 (95% CI 0.722 to 0.815), respectively, with a mean difference of 0.027 (95% CI 0.020 to 0.034; I^2=0%), based on six studies (SOE: Moderate).
  - One study found no clear differences between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in accuracy for identifying people undergoing craniotomy (sensitivity 63%, 95% CI 38 to 84 vs. 68%, 95% CI 43 to 87; and specificity 82%, 95% CI 80 to 84 vs. 83%, 95% CI 81 to 85) (SOE: Low).
  - For the tGCS versus the SMS, the pooled AUROC was 0.809 (95% CI 0.766 to 0.853) versus 0.769 (95% CI 0.711 to 0.827), respectively, with a mean difference of 0.032 (95% CI 0.025 to 0.039, I^2=0%), based on five studies (SOE: Moderate).
  - One study found the out-of-hospital tGCS (cutoff of ≤13) associated with higher sensitivity than the SMS (cutoff of ≤1) for identifying patients who underwent

neurosurgical intervention (60%, 95% CI 56 to 63 vs. 53%, 95% CI 49 to 56) and slightly lower specificity (85%, 95% CI 84 to 85 vs. 86%, 95% CI 86 to 87) (SOE: Low).
    - For the mGCS versus the SMS, the mean difference in the pooled AUROC was 0.002 (95% CI -0.005 to 0.010, I^2=0%), based on four studies (SOE: Moderate).
- Severe brain injury
    - For the tGCS versus the mGCS, the pooled AUROC was 0.791 (95% CI 0.734 to 0.827) versus 0.720 (95% CI 0.666 to 0.774), respectively, with a mean difference of 0.050 (95% CI 0.034 to 0.065; I^2=57%), based on five studies (SOE: Moderate).
    - One study found no difference between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in sensitivity (62%, 95% CI 55% to 68% vs. 61%, 95% CI 54% to 67%) or specificity (85%, 95% CI 83 to 88 vs. 89%, 95% CI 88 to 91) for identifying people with severe head injury (defined as head AIS score of ≥4) (SOE: Low).
    - For the tGCS versus the SMS, the pooled AUROC was 0.763 (95% CI 0.710 to 0.815) versus 0.713 (95% CI 0.654 to 0.771), respectively, with a mean difference of 0.048 (95% CI 0.038 to 0.059, I^2=72%), based on five studies (SOE: Moderate).
    - One study found out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for severe brain injury based on presence of head CT imaging findings (45%, 95% CI 44 to 46 vs. 41%, 95% CI 40 to 42) and similar specificity (89%, 95% CI 89 to 90 vs. 90%, 95% CI 90 to 91) (SOE: Low).
    - For the mGCS versus the SMS, there was no difference in the AUROC (mean difference 0.000, 95% CI -0.008 to 0.007, I^2=0%), based on four studies (SOE: Moderate).
- Emergency intubation
    - For the tGCS versus the mGCS, the pooled AUROC was 0.851 (95% CI 0.794 to 0.908) versus 0.807 (95% CI 0.735 to 0.880), respectively, with a mean difference of 0.038 (95% CI 0.023 to 0.052; I^2=72%), based on five studies (SOE: Moderate).
    - For the tGCS versus the SMS, the pooled AUROC was 0.843 (95% CI 0.823 to 0.864) versus 0.783 (95% CI 0.747 to 0.819), respectively, with a mean difference of 0.040 (95% CI 0.030 to 0.050, I^2=55%), based on five studies (SOE: Moderate).
    - One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for identifying people who underwent emergency intubation (76%, 95% CI 74 to 77 vs. 73%, 95% CI 71 to 74) and slightly lower specificity (89%, 95% CI 89 to 89 vs. 91%, 95% CI 90 to 91) (SOE: Low).
    - For the mGCS versus the SMS, there was no difference in the AUROC (mean difference 0.000, 95% CI -0.007 to 0.007, I^2=0%), based on four studies (SOE: Moderate).
- Trauma center need
    - One study that utilized National Trauma Data Bank (NTDB) data (n=811,143) found small differences between the tGCS versus the mGCS in the AUROC (0.62 vs. 0.61), sensitivity (30% vs. 27%), and specificity (93% vs. 95%) for need of trauma center care (defined as ISS score of >15, intensive care unit [ICU] admission >24 hours, need for urgent surgery, or death in the ED) (SOE: Low).

- Severe injury
  - One study (n=104,035) of children with traumatic brain injury (TBI) in the NTDB found the tGCS was better able to discriminate those with major injury (defined as an ISS score of >15) from those without major injury (AUROC 0.720, 95% CI 0.715 to 0.724 vs. 0.681, 95% CI 0.677 to 0.686) (SOE: Low).

Detailed Synthesis

Eighteen studies directly compared the predictive utility of the tGCS versus the mGCS (12 studies) and/or the SMS (6 studies, Table 1, Appendix H).[17-20,36,48,52-63] All studies were retrospective analyses in which the mGCS or SMS scores were taken from the tGCS (i.e., the tGCS and mGCS or SMS were not assessed independently). Sample sizes ranged from 96 to 811,143. Fifteen studies were conducted in the United States, two studies in Europe, and one in Canada. Four studies restricted enrollment to children;[52,56,61,63] the other studies enrolled adults or mixed populations of adults and children. Four studies utilized data collected started in or after 2006. GCS scores were obtained during out-of-hospital assessment in ten studies and in the ED in four studies; the assessment setting was mixed or unclear in four studies. Four studies focused on patients with TBI[52,57,58,63] and the remainder evaluated mixed trauma populations. Among studies that enrolled mixed trauma patients, the proportion of trauma patients with TBI in studies that reported this information ranged from 5 to 18 percent; none of these studies reported results in subgroup of patients with TBI. No study reported the proportion of intoxicated patients. In two studies, the proportion of patients who underwent out-of-hospital intubation was 0.3 percent[53] and 3.5 percent,[19] it was unclear when the GCS was assessed in intubated patients. Thirteen studies were rated moderate risk of bias and five studies were rated low risk of bias. Eight studies did not report attrition and seven studies reported missing data in more than 20 percent of patients. Most studies on predictive utility focused on measures of discrimination or diagnostic accuracy; adjustment for confounders is generally not performed for either of these measures.

Table 1. Characteristics of head-to-head studies

Author, Year	Settings Years of Study	Assessment Timing Measures and/or Scores Compared	N	Population Characteristics
Acker, *et al.,* 2014[52]	USA, Colorado Urban 2 Level 1 pediatric trauma centers 2002 to 2011	ED A: tGCS B: mGCS	2,231	Age (mean, years): 6.9 (SD 5.8) Male: 65% Race: NR TBI: 100% ISS (median): 17 (IQR: 10-26)
Al-Salamah, *et al.,* 2004[53]	Canada, Ontario Trauma registry 72% urban, 28% suburban or rural 1994 to 2002	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	795	Age (mean, years): 44 (SD 21) Male: 70% Race: NR TBI: NR ISS: NR
Beskind, *et al.,* 2014[17]	USA, Southern Arizona Urban, University Health Network Level 1 trauma center 2008 to 2010	Out-of-hospital A: tGCS B: mGCS	9,816	Age (median, years): 32 (IQR: 20-51) Male: 65.5% Race: NR TBI: NR ISS ≥16: 11.7%
Brown, *et al.,* 2014[54]	USA Trauma registry* 2007 to 2008	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	811,143	Age (median): 39 (IQR: 23-57) Male: 66% Race: NR TBI: NR ISS (median): 9 (IQR: 4-13)
Caterino and Raubenolt, 2012[55]	USA, Ohio Urban, hospitals Trauma and non-trauma centers 2002 to 2007	Out-of-hospital A: tGCS ≤13 B: SMS ≤1	52,412	Age (mean, years): 53 Male: 55.9% White: 79.9% Black: 13.5% Hispanic: 1.5% Other race: 1.7% Race not documented: 3.4% TBI: 15.2% ISS (median): 9 ISS >15: 26.6%
Cicero and Cross, 2013[56]	USA Trauma registry* 2007 to 2009	Out-of-hospital A: tGCS B: mGCS	104,035	Age (mean, years): 12.6 (SD 5.5) Male: 67% Nonwhite race: 38% TBI: NR ISS (mean): 9.9 (SD 10.3) Major injury (ISS >15): 15%

Author, Year	Settings Years of Study	Assessment Timing Measures and/or Scores Compared	N	Population Characteristics
Corrigan, *et al.,* 2014[57]	USA Trauma registry* 2007 to 2010	Out-of-hospital A: tGCS B: mGCS	77,470	NR
Davis, *et al.,* 2006[58]	USA, California (San Diego) Urban, other data NR Date NR	Out-of-hospital and ED A: tGCS B: mGCS	12,882	NR
Eken, *et al.,* 2009[59]	Turkey Tertiary care ED of hospital Level IV trauma center 2006	ED A: tGCS B: mGCS	185	Age (median, years): 59 (range: 18-97) Male: 64% Race: NR TBI: NR ISS: NR
Gill, *et al.,* 2005[20]	USA, California (Loma Linda) Urban, University Level 1 trauma center and children's hospital 1990 to 2002	ED A: tGCS B: mGCS C: SMS	8,412	Age (median, years): 24 (IQR: 15-38) Male: 71.5% Race: NR TBI: 17.1% ISS: NR
Gill, *et al.,* 2006[36]	USA, California (Loma Linda) Urban, University Level 1 trauma center and children's hospital 1990 to 2002	Out-of-hospital A: tGCS B: mGCS C: SMS	7,233	Age (median, years): 24 (IQR: 16-38) Male: 70% Race: NR TBI: 17% ISS: NR
Haukoos, *et al.,* 2007[60]	USA, Colorado Urban, Denver Health Medical Center Level 1 trauma center 1995 to 2004	ED A: tGCS B: mGCS C: SMS	21,170	Age (median, years): 32 (IQR: 21-45) Male: 71% Race: NR TBI: 14% ISS score (median): 9 (IQR: 2-14)
Healey, *et al.,* 2003[18]	USA Trauma registry* 1994 to 2001	Out-of-hospital A: tGCS B: mGCS	202,255	NR
Holmes, *et al.,* 2005[61]	USA, California (Davis) Level 1 trauma center 1998 to 2001	ED A: tGCS B: mGCS	2,043	Ages ≤2 years: 16% Ages >2 years: 84% Male: NR Race: NR TBI: 5% ISS: NR

Author, Year	Settings Years of Study	Assessment Timing Measures and/or Scores Compared	N	Population Characteristics
Ross, *et al.,* 1998[19]	USA, New Jersey Level 1 trauma center 1994 to 1996	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	1,410	Age (mean, years): 37.1 (range: 13-95) Male: 69% Race: NR TBI: NR ISS (mean): 14.4 ISS (median): 13
Thompson, *et al.,* 2011[48]	USA, Colorado Urban, Denver Health Medical Center Level 1 trauma center 1999 to 2008	Out-of-hospital A: tGCS ≤13 B: mGCS score ≤5 B: SMS ≤1	19,408	Age (median, years): 33 (IQR: 22-48) Male: 71% Race: NR TBI: 18% ISS (median): 9 (IQR: 4-17)
Van de Voorde, *et al.,* 2008[63]	Belgium Pediatric trauma registry (PENTA) 2005	Out-of-hospital and ED A: tGCS score ≤13 B: mGCS score ≤5	96	Age (mean, years): 8.2 (SD 5.3) Male: 59% Race: NR TBI: NR ISS (median): 16

ED=emergency department; IQR=interquartile range; ISS=Injury Severity Score; mGCS=motor Glasgow Coma Scale; n=number; NR=not reported; NTDB=National Trauma Data Bank; PENTA=pediatric trauma registry; SD=standard deviation; SMS=Simplified Motor Scale; TBI=traumatic brain injury; tGCS=total Glasgow Coma Scale; USA=United States of America

*Patients from the NTDB data set

Four studies were based on analyses of the NTDB database,[18,54,56,57] but evaluated different populations or outcomes. Sample sizes ranged from 77,470 to 811,143. One of the NTDB studies focused on children,[56] one focused on adults,[57] and two evaluated mixed populations.[18,54] There were also two studies[20,36] that used data from a trauma center in in Loma Linda, California and three studies[48,52,60] that used data from the Denver area trauma system in which there could be some overlap in the populations assessed. In sensitivity analyses, we excluded two studies[20,60] from these trauma systems that focused on GCS scores obtained in the ED, since other studies[36,48] from these trauma systems evaluated out-of-hospital GCS scores around the same time period.

The most commonly evaluated outcome was in-hospital mortality. Other outcomes reported in at least five studies were severe brain injury, receipt of neurosurgical intervention, and intubation (Tables 2). The proportion of patients who experienced in-hospital mortality ranged from 3 percent to 18 percent,[17,19,20,36,48,52-56,59,60,62,63] the proportion with severe brain injury (defined in one study[61] as computed tomography [CT] imaging findings of skull fracture, contusion or hemorrhage or acute intervention for TBI [neurosurgical procedure, hospitalization >2 days, antiepileptic medications for >7 days]; the others defined severe brain injury based on CT imaging alone[20,36,48,55,60]) ranged from 5 percent to 39 percent, the proportion who underwent a neurosurgical intervention (defined as craniotomy in one study[52] and as a composite outcome including various neurosurgical procedures, ventriculostomy, and/or intracerebral pressure monitoring in the others[17,20,36,48,55,60]) ranged from 1.5 percent to 10 percent; and the proportion who were intubated (out-of-hospital, ED, or both) ranged from 4 percent to 26 percent.[17,20,36,48,53,55,60]

Table 2. Proportion of patients experiencing outcomes in head-to-head studies on predictive utility

Study, Year	In-Hospital Mortality	Neurosurgical Intervention	Severe Brain Injury	Severe Injury	Intubation
Acker, *et al.,* 2014[52]	8.4%	10.4%*	--	--	--
Al-Salamah, *et al.,* 2004[53]	18%	--	--	--	16%†
Beskind, *et al.,* 2014[17]	2.9%	3.8%	--	--	4.1%‡
Brown, *et al.,* 2014[§, 54]	4.3%	--	--	39%‖	--
Caterino, and Raubenolt, 2012[55]	5.8%	1.5%	15%¶	--	7.6%‡
Cicero, and Cross, 2013[§, 56]	3.8%	--	--	21**	--
Corrigan, *et al.,* 2014[§, 57]	--	--	--	--	--
Davis, *et al.,* 2006[58]	--	--	--	--	--
Eken, *et al.,* 2009[59]	14%	--	--	--	--
Gill, *et al.,* 2005[20]	11%	9.3%	17%¶	--	26%†
Gill, *et al.,* 2006[36]	10%	8.8%	17%¶	--	26%†
Haukoos, *et al.,* 2007[60]	5.5%	6.6%	14%¶	--	18%‡
Healey, *et al.,* 2003[§, 18]	6%	--	--	--	--
Holmes, *et al.,* 2005[61]	--	--	6.3%††	--	--
Ross, *et al.,* 1998[19]	6.6%	--	--	--	3.5%‡‡
Thompson, *et al.,* 2011[48]	5.8%	7.8%	18%¶	--	18%‡
Timmons, *et al.,* 2011[62]	15%	--	--	--	--
Van de Voorde, *et al.,* 2008[63]	11%	--	--	--	--

*Craniotomy only
†Intubation in emergency department
‡Intubation in pre-hospital setting or emergency department
§Studies from NTDB database
‖Injury Severity Score >15, intensive care unit admission ≥24 hours, need for urgent surgery (emergency department disposition to the operating room), or death in the emergency department
¶Skull fracture with underlying brain injury, intracranial hemorrhage, cerebral contusion, or non-specific intracranial injury
**Injury Severity Score >15
††Traumatic brain injury on computed tomography scan (intracranial hemorrhage, hematoma, contusion, or cerebral edema) or in need of acute intervention (neurosurgical procedure, antiepileptic medication for >7 days, neurologic deficit persisting until discharge, or ≥2 nights of hospitalization for treatment for blunt head injury)
‡‡Intubation in out-of-hospital setting

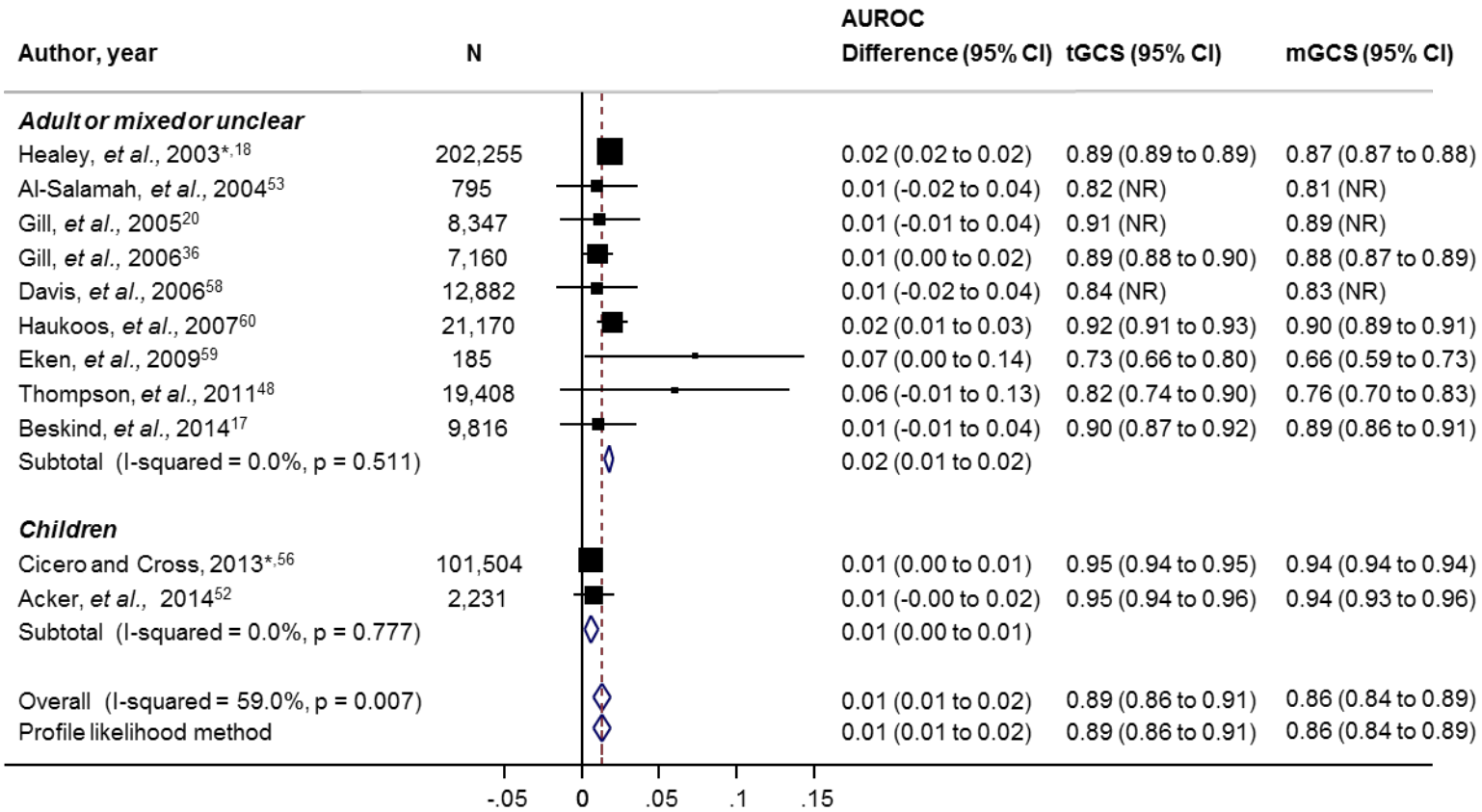
Mortality

Total Glasgow Coma Scale Versus Motor Glasgow Coma Scale

The tGCS was slightly better than the mGCS at discriminating individuals who experienced in-hospital mortality from those who survived to hospital discharge (Table 3). Based on 11 studies, the pooled AUROC for the tGCS was 0.886 (95% CI 0.863 to 0.908) and for the mGCS

was 0.864 (95% CI 0.839 to 0.890), with a pooled mean difference of 0.013 (95% CI 0.007 to 0.019; I^2=59%; Figure 3).[17,18,20,36,48,52,53,56,58-60] Results were unchanged when the analysis was performed using the profile likelihood method. Stratification of studies according to whether they focused on children (2 studies, mean difference in AUROC 0.006, 95% CI 0.002 to 0.011, I^2=0%)[52,56] or evaluated adults or mixed populations (9 studies, mean difference in AUROC 0.017, 95% CI 0.015 to 0.020, I^2=0%)[17,18,20,36,48,53,58-60] eliminated statistical heterogeneity, though the pooled mean difference for each subgroup was very similar (Table 4). Results were also similar in subgroup analyses stratified according to use of out-of-hospital (6 studies, mean difference in AUROC 0.012, 95% CI 0.004 to 0.020, I^2=74%)[17,18,36,48,56,58] or ED GCS scores (3 studies, mean difference in AUROC 0.020, 95% CI 0.006 to 0.034, I^2=20%)[20,59,60] or years in which data were collected (all data after 2006 vs. some or all before 2006), or when analyses were restricted to low risk of bias studies,[17,18,20,36,60] studies of patients with TBI,[17,52,58] or studies conducted in the United States (Table 4). The two largest studies (n=202,255 and 101,504, compared with 185 to 21,170 in the other studies) were based on the NTDB database.[18,56] Estimates from the NTDB studies were very similar (differences in the AUROC 0.02, 95% CI 0.02 to 0.02[18] and 0.01, 95% CI 0.00 to 0.01[56]) and excluding the NTDB studies had little effect on estimates (Table 4). When multiple studies were available from a trauma center, restricting the analysis to the most recent study from each trauma center that used out-of-hospital GCS scores (excluding 3 studies[20,52,60]) also had little effect on estimates.[20,48]

Figure 3. Pooled AUROC of mortality for the total Glasgow Coma Scale versus the motor component only



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; mGCS=motor Glasgow Coma Scale; n=number; NR=not reported; tGCS=total Glasgow Coma Scale

*Patients from the NTDB data set

Table 3. Summary of discrimination (AUROC) of head-to-head studies

Author, Year	Assessment Timing Measures and/or Scores Compared	N	Age	In-Hospital Mortality (95% CI)	Neurosurgical Intervention (95% CI)	Severe Brain Injury (95% CI)	Intubation (95% CI)
Acker, *et al.,* 2014[52]	ED A: tGCS B: mGCS	2,231	≤18 years Mean: 6.9 years	0.949 (0.938 to 0.961) vs. 0.941 (0.926 to 0.957), p=0.06	0.642 (0.603 to 0.681) vs. 0.638 (0.601 to 0.675), p=0.64* 0.808 (0.784 to 0.832) vs. 0.774 (0.748 to 0.800), p<0.001†	NR	NR
Al-Salamah, *et al.,* 2004[53]	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	795	≥16 years Mean: 44 years	0.82 vs. 0.81, p=NR	NR	NR	NR
Beskind, *et al.,* 2014[17]	Out-of-hospital A: tGCS B: mGCS	9,816	Mean: 32 years	0.899 (0.874 to 0.923) vs. 0.888 (0.864 to 0.913), mean difference=0.010 (0.002 to 0.018)	0.571 (0.533 to 0.609) vs. 0.570 (0.531 to 0.608), mean difference=0.002 (-0.013 to 0.016)	NR	0.966 (0.955 to 0.976) vs. 0.948 (0.933 to 0.963), mean difference=0.018 (0.011 to 0.024)
Caterino and Raubenolt, 2012[55]	Out-of-hospital A: tGCS ≤13 B: SMS ≤1	52,412	≥16 years Mean: 53 years	0.85 (0.84 to 0.86) vs. 0.82 (0.81 to 0.83)	0.75 (0.73 to 0.77) vs. 0.70 (0.68 to 0.72)	0.72 (0.71 to 0.72) vs. 0.66 (0.65 to 0.66)	0.86 (0.85 to 0.87) vs. 0.83 (0.82 to 0.83)
Cicero and Cross, 2013‡, [56]	Out-of-hospital A: tGCS B: mGCS	104,035	<19 years Mean: 12.6 years	0.946 (0.941 to 0.951) vs. 0.940 (0.935 to 0.945)	NR	NR	NR
Corrigan, *et al.,* 2014‡, [57]	Out-of-hospital A: tGCS B: mGCS	77,470	≥18 years	0.886 (NR) vs. 0.878 (NR)	NR	NR	NR
Davis, *et al.,* 2006[58]	Out-of-hospital and ED A: tGCS B: mGCS	12,882	NR	0.84 (NR) vs. 0.83 (NR)	0.80 (NR) vs. 0.78 (NR)	NR	NR
Eken, *et al.,* 2009[59]	ED A: tGCS B: mGCS	185	>17 years	0.735 (0.655 to 0.797) vs. 0.662 (0.589 to 0.730)	NR	NR	NR
Gill, *et al.,* 2005[20]	ED A: tGCS B: mGCS C: SMS	8,412	Median of 24 years	0.906 (NR) vs. 0.894 (NR) vs. 0.878 (NR)	0.874 (NR) vs. 0.848 (NR) vs. 0.851 (NR)	0.826 (NR) vs. 0.789 (NR) vs. 0.791 (NR)	0.865 (NR) vs. 0.826 (NR) vs. 0.826 (NR)

Author, Year	Assessment Timing Measures and/or Scores Compared	N	Age	In-Hospital Mortality (95% CI)	Neurosurgical Intervention (95% CI)	Severe Brain Injury (95% CI)	Intubation (95% CI)
Gill, *et al.,* 2006[36]	Out-of-hospital A: tGCS B: mGCS C: SMS	7,233	Median of 24 years	0.89 (0.88 to 0.90) vs. 0.88 (0.87 to 0.89) vs. 0.86 (0.86 to 0.89)	0.86 (0.85 to 0.88) vs. 0.84 (0.82 to 0.85) vs. 0.83 (0.81 to 0.84)	0.83 (0.82 to 0.84) vs. 0.79 (0.78 to 0.81) vs. 0.79 (0.77 to 0.80)	0.83 (0.81 to 0.84) vs. 0.79 (0.78 to 0.80) vs. 0.79 (0.77 to 0.80)
Haukoos, *et al.,* 2007[‡, 60]	ED A: tGCS B: mGCS C: SMS	21,170	Median of 32 years	0.92 (0.91 to 0.93) vs. 0.90 (0.89 to 0.91) vs. 0.89 (0.88 to 0.90)	0.83 (0.82 to 0.84) vs. 0.80 (0.79 to 0.81) vs. 0.80 (0.79 to 0.81)	0.76 (0.75 to 0.77) vs. 0.71 (0.70 to 0.72) vs. 0.71 (0.70 to 0.72)	0.86 (0.85 to 0.87) vs. 0.81 (0.80 to 0.82) vs. 0.81 (0.80 to 0.82)
Healey, *et al.,* 2003[18]	Out-of-hospital A: tGCS B: mGCS	202,255	NR	0.891 (0.888 to 0.894) vs. 0.873 (0.870 to 0.875), p=0.000	NR	NR	NR
Holmes, *et al.,* 2005[61]	ED A: tGCS B: mGCS	2,043	≤2 years	NR	NR	Ages ≤2 years: 0.72 (0.56 to 0.87) vs. 0.60 (0.48 to 0.72) Ages >2 years: 0.82 (0.76 to 0.87) vs. 0.71 (0.65 to 0.77) AUC (95% CI) for TBI in need of acute intervention Ages ≤2 years: 0.97 (0.94 to 1.0) vs. 0.76 (0.59 to 0.93) Ages >2 years 0.87 (0.83 to 0.92) vs. 0.76 (0.71 to 0.81)	NR
Thompson, *et al.,* 2011[48]	Out-of-hospital A: tGCS ≤13 B: mGCS score ≤5 B: SMS ≤1	19,408	All Median: 33 years	0.82 (0.74 to 0.90) vs. 0.76 (0.70 to 0.83) vs. 0.74 (0.70 to 0.77)	0.70 (0.64 to 0.77) vs. 0.66 (0.61 to 0.71) vs. 0.66 (0.64 to 0.69)	0.66 (0.60 to 0.71) vs. 0.61 (0.57 to 0.65) vs. 0.61 (0.58 to 0.64)	0.70 (0.63 to 0.77) vs. 0.65 (0.60 to 0.70) vs. 0.65 (0.62 to 0.67)

AUC=area under the receiver operating characteristics curve; AUCROC=area under the receiver operating characteristics curve; CI=confidence interval; ED=emergency department; mGCS=motor Glasgow Coma Scale; n=number; NR=not reported; SMS=Simplified Motor Scale; TBI=traumatic brain injury; tGCS=total Glasgow Coma Scale; vs.=versus

*Craniotomy only
†Intracranial pressure monitoring only
‡Studies from NTDB database

Table 4. Pooled AUROC results of head-to-head studies

Outcome and Analysis	tGCS vs. mGCS, Difference in AUROC (95% CI)	Number of Studies	I^2	tGCS vs. SMS, Difference in AUROC (95% CI)	Number of Studies	I^2
Mortality, overall	0.013 (0.007 to 0.019)	11	59%	0.030 (0.024 to 0.036)	5	0%
Adults or mixed	0.017 (0.015 to 0.020)	9	0%	0.030 (0.024 to 0.036)	5	0%
Children	0.006 (0.002 to 0.011)	2	0%	--	--	--
Excluding NTDB studies	0.014 (0.008 to 0.019)	9	0%	0.030 (0.024 to 0.036)	5	0%
Excluding studies with potential overlap*	0.013 (0.005 to 0.020)	8	68%	0.031 (0.023 to 0.039)	3	0%
Out-of-hospital GCS	0.012 (0.004 to 0.020)	6	74%	0.031 (0.023 to 0.039)	3	0%
ED GCS	0.020 (0.006 to 0.034)	3	20%	0.030 (0.020 to 0.039)	2	0%
U.S. setting	0.013 (0.007 to 0.019)	9	63%	0.030 (0.024 to 0.036)	5	0%
TBI patients	0.009 (-0.002 to 0.020)	3	0%	--	--	--
Low risk of bias studies	0.017 (0.015 to 0.020)	5	0%	0.030 (0.022 to 0.037)	3	0%
Enrollment before 2006	0.016 (0.012 to 0.020)	9	11%	0.030 (0.024 to 0.036)	5	0%
Enrollment after 2006	0.006 (0.001 to 0.011)	2	0%	--	--	--
Neurosurgical intervention, overall	0.027 (0.020 to 0.034)	6	0%	0.032 (0.025 to 0.039)	5	0%
Adults or mixed	0.026 (0.019 to 0.034)	5	0%	0.032 (0.025 to 0.039)	5	0%
Children	0.034 (0.009 to 0.059)	1	--	--	--	--
Excluding studies with potential overlap*	0.021 (0.008 to 0.034)	3	0%	0.038 (0.024 to 0.052)	3	19%
Out-of-hospital GCS	0.021 (0.008 to 0.034)	3	0%	0.038 (0.024 to 0.052)	3	19%
ED GCS	0.030 (0.020 to 0.039)	2	0%	0.029 (0.020 to 0.038)	2	0%
U.S. setting	0.027 (0.020 to 0.034)	6	0%	0.032 (0.025 to 0.039)	5	0%
TBI patients	0.017 (-0.022 to 0.056)	2	66%	--	--	--
Low risk of bias studies	0.026 (0.019 to 0.034)	4	0%	0.029 (0.021 to 0.037)	3	0%
Enrollment before 2006	0.028 (0.020 to 0.035)	5	0%	0.032 (0.025 to 0.039)	5	0%
Enrollment after 2006	0.019 (-0.009 to 0.047)	1	--	--	--	--
Severe brain injury, overall	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Adults or mixed	0.046 (0.038 to 0.054)	4	0%	0.048 (0.038 to 0.059)	5	72%
Children	0.121 (0.068 to 0.174)	1	--	--	--	--
Excluding NTDB studies	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Excluding studies with potential overlap*	0.065 (0.020 to 0.111)	3	76%	0.051 (0.034 to 0.068)	3	74%
Out-of-hospital GCS	0.041 (0.028 to 0.053)	2	0%	0.051 (0.034 to 0.068)	3	74%
ED GCS	0.060 (0.028 to 0.093)	3	73%	0.044 (0.030 to 0.059)	2	51%
U.S. setting	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
TBI patients	--	--	--	--	--	--

Outcome and Analysis	tGCS vs. mGCS, Difference in AUROC (95% CI)	Number of Studies	I^2	tGCS vs. SMS, Difference in AUROC (95% CI)	Number of Studies	I^2
Low risk of bias studies	0.046 (0.038 to 0.053)	3	0%	0.044 (0.035 to 0.053)	3	25%
Enrollment before 2006	0.050 (0.034 to 0.065)	5	57%	0.048 (0.038 to 0.059)	5	72%
Enrollment after 2006	--	--	--	--	--	--
Emergency intubation, overall	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
Adults or mixed	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
Children	--	--	--	--	--	--
Excluding studies with potential overlap*	0.031 (0.012 to 0.050	3	65%	0.033 (0.025 to 0.040)	3	0%
Out-of-hospital GCS	0.031 (0.012 to 0.050)	3	65%	0.033 (0.025 to 0.040)	3	0%
ED GCS	0.048 (0.039 to 0.058)	2	0%	0.048 (0.039 to 0.057)	2	0%
U.S. setting	0.038 (0.023 to 0.052)	5	72%	0.040 (0.030 to 0.050)	5	55%
TBI patients	0.011 (-0.010 to 0.032)	1	--	--	--	--
Low risk of bias studies	0.037 (0.022 to 0.052)	4	79%	0.046 (0.038 to 0.054)	3	0%
Enrollment before 2006	0.046 (0.038 to 0.053)	4	0%	0.040 (0.030 to 0.050)	5	55%
Enrollment after 2006	0.018 (0.005 to 0.031)	1	--	--	--	--

AUROC=area under the receiver operating characteristics curve; CI=confidence interval; ED=emergency department; GCS=Glasgow Coma Scale; mGCS=motor Glasgow Coma Scale; NTDB=National Trauma Data Bank; SMS=Simplified Motor Score; TBI=traumatic brain injury; tGCS=total Glasgow Coma Scale; U.S.=United States of America

*When multiple studies published from the same trauma center, analysis restricted to the most recent study using pre-hospital GCS scores (excluded Gill 2005,[20] Haukoos 2007,[60] Acker 2014[52])

Data on the diagnostic accuracy of the tGCS versus the mGCS were limited (Table 5). In four studies, sensitivity ranged from 71 percent to 100 percent for the tGCS (cutoff of ≤13) and from 72 percent to 100 percent for the mGCS (cutoff of ≤5); differences in sensitivity ranged from 0 percent to 1 percent (Table 5).[19,53,54,63] Specificity ranged from 68 percent to 85 percent for the tGCS and from 73 percent to 86 percent for the mGCS; difference in sensitivity ranged from 0 percent to 5 percent in favor of the mGCS, though the CIs overlapped in each study. Two of the studies were conducted in mixed populations of adults and children and one of the studies focused on children (sensitivity 100%, specificity 74%).[63] The latter study found that the specificity of the tGCS decreased at higher cutoffs (74% for a cutoff of ≤13, 71% for ≤14 and 56% for ≤15) with little change in sensitivity;[63] the other studies did not report specificity at tGCS cutoffs other than 13 or less. One study found that calibration of the tGCS and mGCS was similarly poor based on the Hosmer-Lemeshow test (p-value <0.01 for both scales).[53]

Table 5. Summary of diagnostic accuracy outcomes for head-to-head studies

Author, year	Assessment Timing Measures and/or Scores Compared	N	Age	In-hospital Mortality
Al-Salamah, *et al.,* 2004[53]	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	795	≥16 years Mean: 44 years	Sensitivity (95% CI)*: 80.28% (72.78 to 86.48) vs. 80.28% (72.78 to 86.48) Specificity (95% CI)*: 67.99% (64.26 to 71.56) vs. 73.05% (69.47 to 76.42) PPV (95% CI)*: 35.29% (30.08 to 40.78) vs. 39.31% (33.65 to 45.19) NPV (95% CI)*: 94.07% (91.54 to 96.02) vs. 94.46% (92.09 to 96.28)
Brown, *et al.,* 2014[†, 54]	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	811,143	≥3 years Median: 39 years	Sensitivity: 30.3% vs. 26.7% Specificity: 93.1% vs. 95.1%
Caterino and Raubenolt, 2012[55]	Out-of-hospital A: tGCS ≤13 B: SMS ≤1	52,412	≥16 years Mean: 53 years	Sensitivity (95% CI)*: 75.03% (73.45 to 76.56) vs. 72.20% (70.57 to 73.79) Specificity (95% CI)*: 87.63% (87.34 to 87.92) vs. 89.42% (89.14 to 89.69) PPV (95% CI)*: 27.20% (26.25 to 28.17) vs. 29.59% (28.55 to 30.64) NPV (95% CI)*: 98.28% (98.15 to 98.40) vs. 98.12% (97.99 to 98.25)
Ross, et al., 1998[19]	Out-of-hospital A: tGCS score ≤13 B: mGCS score ≤5	1,410	≥13 yearsMean: 37 years	Sensitivity (95% CI)*: 71.28% (61.02 to 80.14) vs. 72.34% (62.15 to 81.07) Specificity (95% CI)*: 84.95% (82.91 to 86.84) vs. 86.02% (84.03 to 87.85) PPV (95% CI)*: 25.28% (20.16 to 30.96) vs. 26.98% (21.61 to 32.91) NPV (95% CI)*: 97.64% (96.59 to 98.44) vs. 97.75% (96.73 to 98.53)
Van de Voorde, et al., 2008[63]	Out-of-hospital and ED A: tGCS score ≤13 B: mGCS score ≤5	96	≤18 years Mean: 8 years	Sensitivity (95% CI): 100% (69.15 to 100) vs. 100% (69.15 to 100) Specificity (95% CI): 74.39% (63.56 to 83.40) vs. 74.36% (63.21 to 83.58)

Author, year	Neurosurgical Intervention	Severe Brain Injury	Intubation
Al-Salamah, *et al.,* 2004[53]	NR	NR	NR
Brown, *et al.,* 2014[54]	NR	NR	NR
Caterino and Raubenolt, 2012[55]	Sensitivity (95% CI)*: 60.05% (56.53 to 63.50) vs. 52.93% (49.37 to 56.46) Specificity (95% CI)*: 84.70% (84.39 to 85.01) vs. 86.40% (86.10 to 86.69) PPV (95% CI)*: 5.64% (5.15 to 6.15) vs. 5.59% (5.08 to 6.14) NPV (95% CI)*: 99.29% (99.20 to 99.36) vs. 99.18% (99.09 to 99.26)	Sensitivity (95% CI)*: 45.40% (44.30 to 46.50) vs. 40.81% (39.72 to 41.89) Specificity (95% CI)*: 89.30% (89.01 to 89.59) vs. 90.50% (90.22 to 90.77) PPV (95% CI)*: 43.20% (42.13 to 44.27) vs. 43.50% (42.38 to 44.64) NPV (95% CI)*: 90.12% (89.84 to 90.40) vs. 89.51% (89.22 to 89.79)	*Any emergency intubation* Sensitivity (95% CI)*: 75.50% (74.13 to 76.83) vs. 72.71% (71.30 to 74.09) Specificity (95% CI)*: 88.90% (88.62 to 89.18) vs. 90.60% (90.34 to 90.86) PPV (95% CI)*: 35.87% (34.84 to 36.91) vs. 38.88% (37.77 to 40.00) NPV (95% CI)*: 97.78% (97.64 to 97.92) vs. 97.58% (97.44 to 97.72) *ED intubation* Sensitivity (95% CI)*: 76.89% (75.43 to 78.31) vs. 74.09% (72.57 to 75.57) Specificity (95% CI)*: 88.20% (87.91 to 88.48) vs. 89.83% (89.56 to 90.09) PPV (95% CI)*: 30.82% (29.83 to 31.82) vs. 33.22% (32.15 to 34.30) NPV (95% CI)*: 98.24% (98.11 to 98.36) vs. 98.07% (97.94 to 98.19)
Ross, *et al.,* 1998[19]	Sensitivity (95% CI)*: 63.16% (38.36 to 83.71) vs. 68.42% (43.45 to 87.42) Specificity (95% CI)*: 81.81% (79.68 to 83.81) vs. 82.82% (80.73 to 84.77) PPV (95% CI)*: 4.53% (2.36 to 7.78) vs. 5.16% (2.78 to 8.66) NPV (95% CI)*: 99.39% (98.74 to 99.75) vs. 99.48% (98.88 to 99.81)	Sensitivity: 61.72% (54.76 to 68.34) vs. 60.77% (53.79 to 67.43) Specificity: 85.47% (83.05 to 87.67) vs. 89.59% (87.73 to 91.26) PLR: 4.25 (3.52 to 5.13) vs. 5.84 (4.79 to 7.12) NLR: 0.45 (0.38 to 0.53) vs. 0.44 (0.37 to 0.52) PPV: 48.68% (42.52 to 54.87) vs. 50.40% (44.05 to 56.73) NPV: 90.91% (88.81 to 92.73) vs. 92.92% (91.29 to 94.33)	NR
Van de Voorde, *et al.,* 2008[63]	NR	NR	NR

CI=confidence interval; ED=emergency department; mGCS=motor Glasgow Coma Scale; n=number; NLR=negative likelihood ratio; NPV=negative predictive value; NR=not reported; PLR=positive likelihood ratio; PPV=positive predictive values; SMS=Simplified Motor Scale; tGCS=total Glasgow Coma Scale; vs.=versus

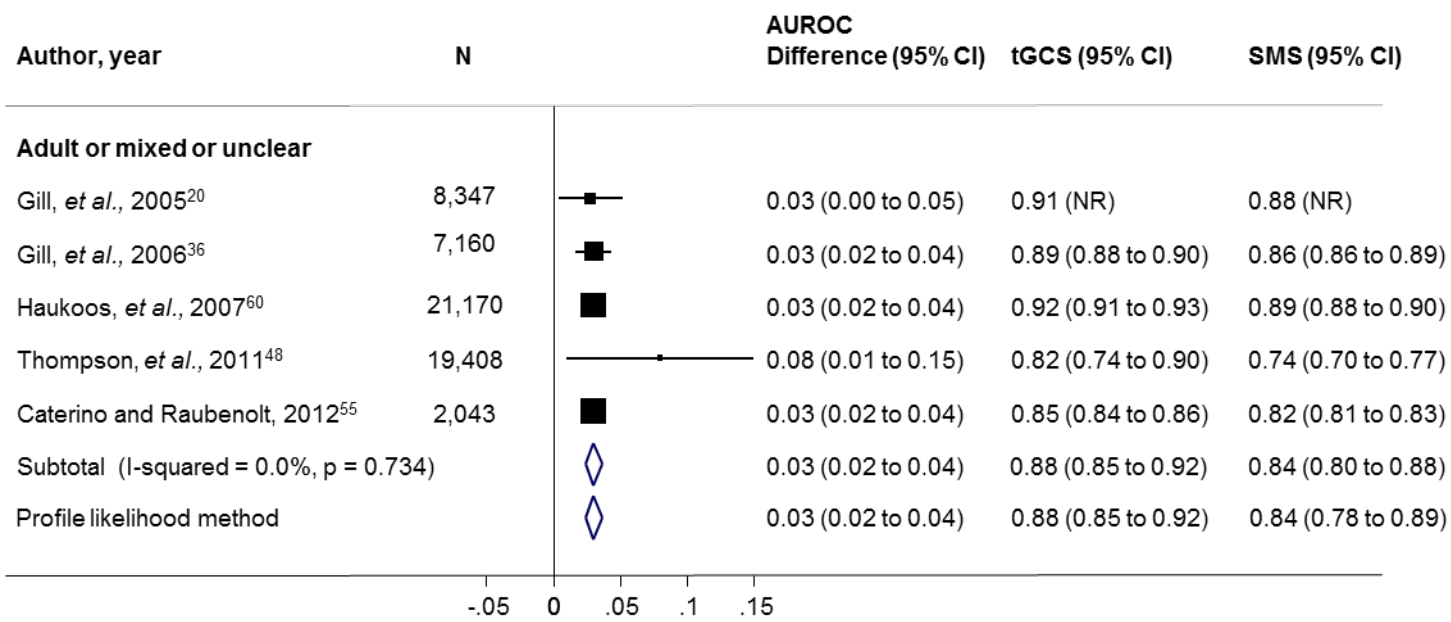
*Calculated
†Study from NTDB database

Total Glasgow Coma Scale Versus Simplified Motor Scale

The tGCS was slightly better than the SMS at discriminating patients who experienced in-hospital mortality from patients who survived to hospital discharge. Based on five studies, the pooled AUROC for the tGCS was 0.884 (95% CI 0.852 to 0.916) and for the SMS was 0.840 (95% CI 0.802 to 0.878), with a mean difference of 0.030 (95% CI 0.024 to 0.036, I^2=0%; Figure 4).[20,36,48,55,60] All of the studies were conducted in the United States and none focused on children or patients with TBI. There were no differences in estimates between studies that utilized out-of-hospital GCS scores or ED GCS scores (Table 4). Results were also unchanged when analyses were restricted to low risk of bias studies, or when studies from the same center with potential overlap[20,60] were excluded. No study was based on the NTDB database.

One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity versus the SMS (cutoff of ≤1) (75% vs. 72%) and slightly lower specificity (88% vs. 89%; Table 5).[55]

Figure 4. Pooled AUROC of mortality for the total Glasgow Coma Scale versus the Simplified Motor Scale



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; n=number; NR=not reported; SMS=Simplified Motor Scale; tGCS=total Glasgow Coma Scale

Motor Glasgow Coma Scale Versus Simplified Motor Scale

The mGCS was slightly better than the SMS at discriminating patients who experienced in-hospital mortality from patients who survived to hospital discharge (4 studies, mean difference in AUROC 0.014, 95% CI 0.006 to 0.021, I^2=0%).[20,36,48,60] There was no statistical heterogeneity and findings were unchanged in sensitivity and subgroup analyses.

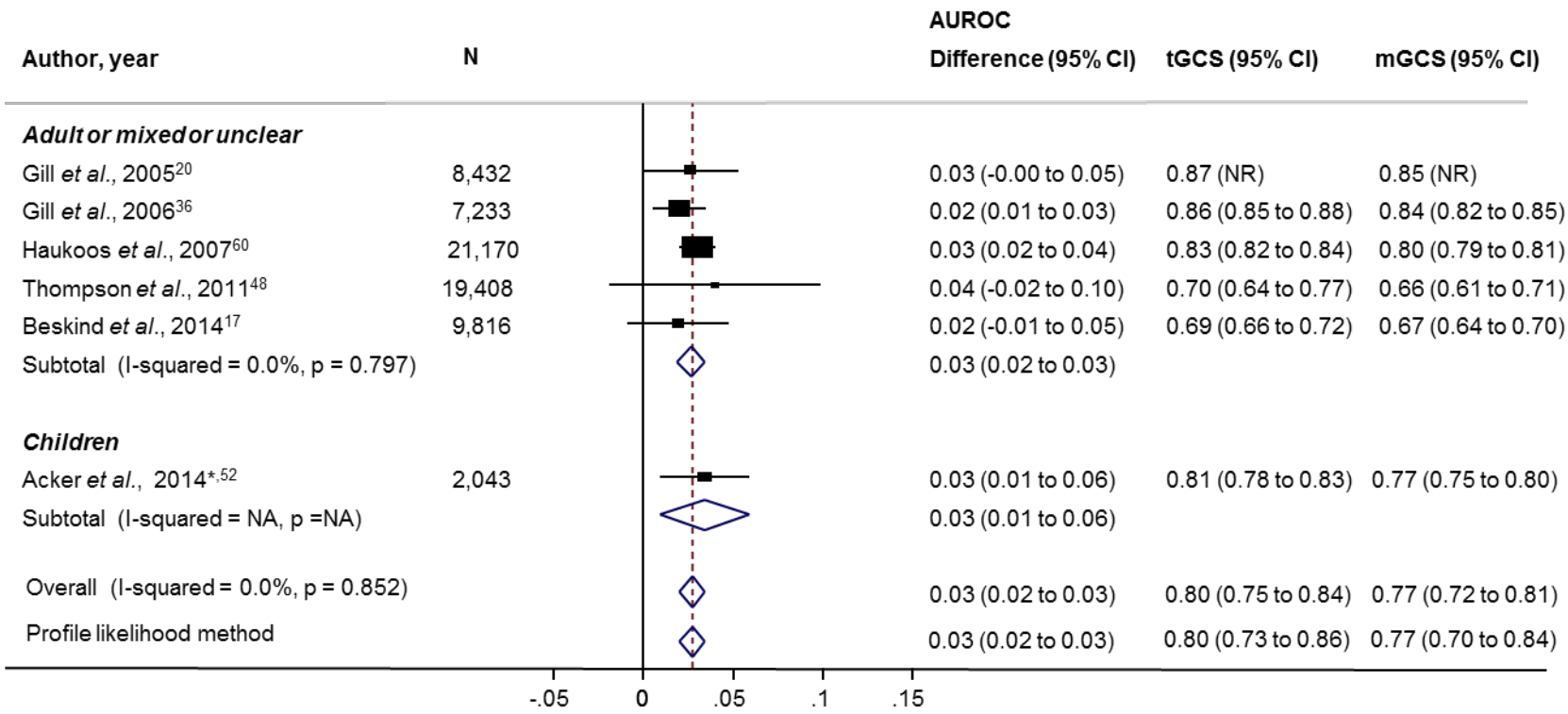
Neurosurgical Intervention

Total Glasgow Coma Scale Versus Motor Glasgow Coma Scale

The tGCS was slightly better than the mGCS at discriminating patients who went on to receive a neurosurgical intervention from those who did not. Based on six studies, the pooled AUROC for the tGCS was 0.798 (95% CI 0.754 to 0.842) and for the mGCS was 0.769 (95% CI 0.722 to 0.815) for a mean difference of 0.027 (95% CI 0.020 to 0.034; I^2=0%; Figure 5).[17,20,36,48,52,60] Results were similar when the analysis was performed using the profile likelihood method. Results were similar in one study[52] that restricted enrollment to children (mean difference in AUROC 0.034, 95% CI 0.009 to 0.059) and the other five studies, which evaluated mixed populations of adults and children (mean difference in AUROC 0.026, 95% CI 0.019 to 0.034, I^2=0%). One of the studies focused on craniotomy only,[52] but reported results similar to studies that evaluated craniotomy plus other neurosurgical interventions. Results were also similar when studies were stratified according to whether they used out-of-hospital or ED GCS scores, or when analyses were restricted to studies conducted in the United States, studies that focused on TBI patients, or low risk of bias studies (Table 4). No study was based on the NTDB database.

One study found no clear differences between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in accuracy for identifying patients undergoing craniotomy (sensitivity 63%, 95% CI 38 to 84 vs. 68%, 95% CI 43 to 87 and specificity 82%, 95% CI 80 to 84 vs. 83%, 95% CI 81 to 85; Table 5).[19]

Figure 5. Pooled AUROC of neurological intervention for the total Glasgow Coma Scale versus the motor component only



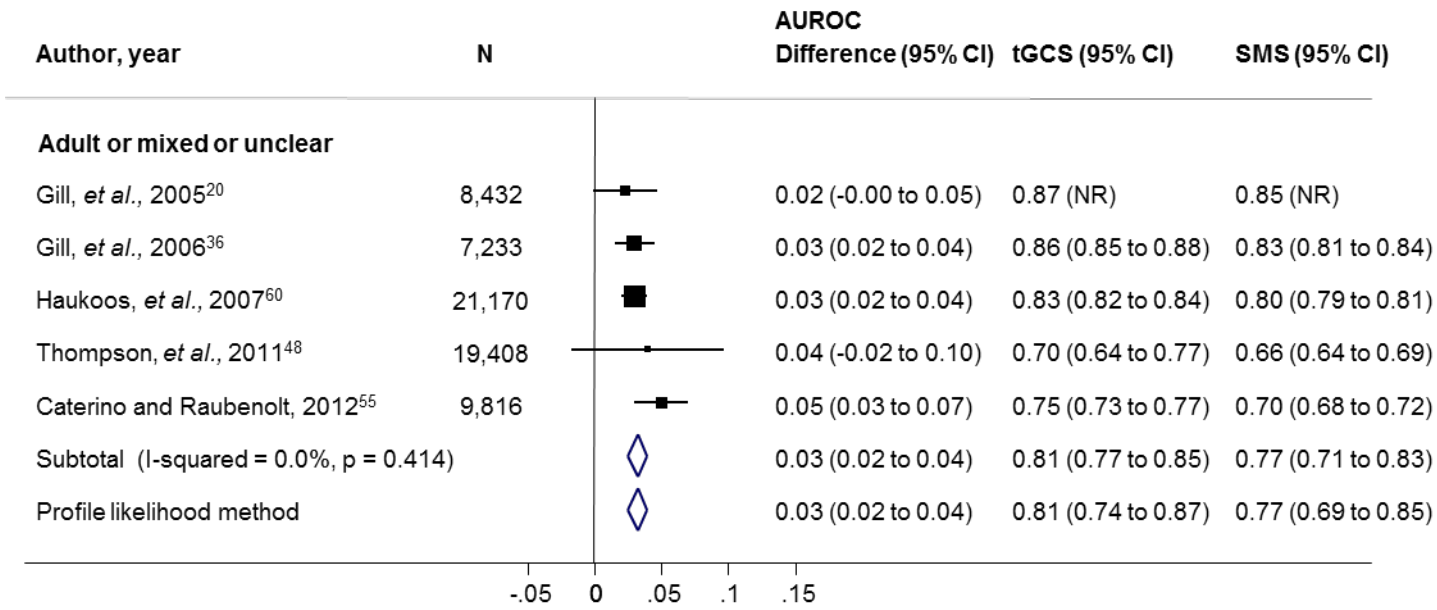
AUROC=area under the receiver operating characteristics curve; CI=confidence interval; mGCS=motor Glasgow Coma Scale; n=number; NA=not applicable; NR=not reported; tGCS=total Glasgow Coma Scale

*Intracranial pressure monitoring only

Total Glasgow Coma Scale Versus Simplified Motor Scale

The tGCS was slightly better than the SMS at discriminating patients who underwent a neurosurgical intervention from patients who did not undergo a neurosurgical intervention; results were very similar to the comparison of tGCS versus mGCS. Based on five studies, the pooled AUROC for the tGCS was 0.809 (95% CI 0.766 to 0.853) and for the mGCS was 0.769 (95% CI 0.711 to 0.827), with a mean difference of 0.032 (95% CI 0.025 to 0.039, I^2=0%; Figure 6).[20,36,48,55,60] All of the studies were conducted in the United States and none focused on children or patients with TBI. There were no differences in estimates between studies that utilized out-of-hospital GCS scores or ED GCS scores (Table 4). Results were also unchanged when analyses were restricted to low risk of bias studies, or when studies with potential overlap[20,60] were excluded. No study utilized data from the NTDB.

Figure 6. Pooled AUROC of neurological intervention for the total Glasgow Coma Scale versus Simplified Motor Scale



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; n=number; NR=not reported; SMS=Simplified Motor Scale; tGCS=total Glasgow Coma Scale

One study found the out-of-hospital tGCS (cutoff of ≤13) associated with higher sensitivity than the SMS (cutoff of ≤1) for identifying patients who underwent neurosurgical intervention (60%, 95% CI 56 to 63 vs. 53%, 95% CI 49 to 56) and slightly lower specificity (85%, 95% CI 84 to 85 vs. 86%, 95% CI 86 to 87; Table 5).[55]

Motor Glasgow Coma Scale Versus Simplified Motor Scale

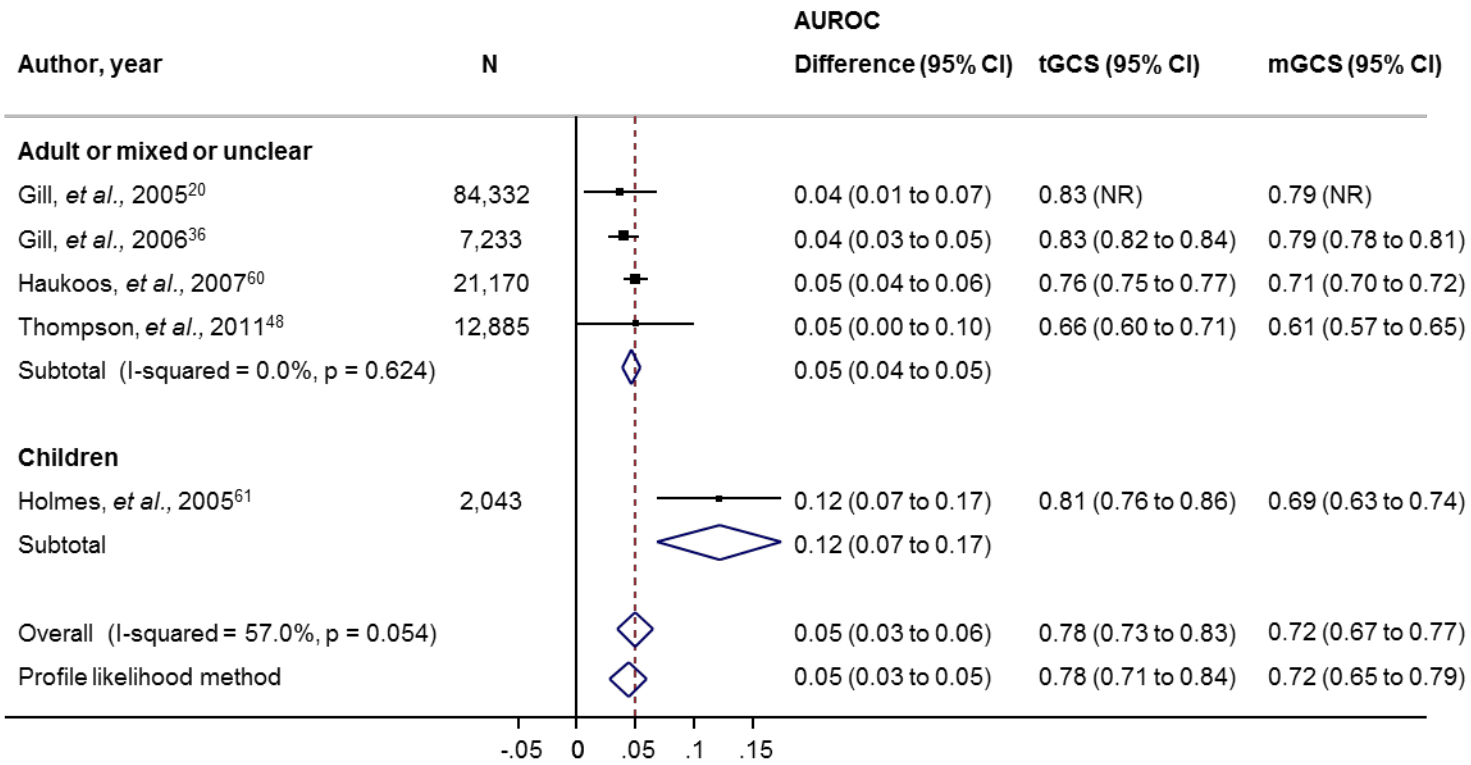
There was no difference between the mGCS versus the SMS in ability to discriminate patients who would undergo a neurosurgical intervention from those who would not undergo a neurosurgical intervention (4 studies, mean difference in AUROC 0.002, 95% CI -0.005 to 0.010, I^2=0%).[20,36,48,60] There was no statistical heterogeneity and findings were unchanged in sensitivity and stratified analyses.

Severe Brain Injury

Total Glasgow Coma Scale Versus Motor Glasgow Coma Scale

The tGCS was slightly better than the mGCS at discriminating patients found to have a severe brain injury from those without severe brain injury. Based on five studies, the pooled AUROC for the tGCS was 0.791 (95% CI 0.734 to 0.827) and for the mGCS was 0.720 (95% CI 0.666 to 0.774), with a mean difference of 0.050 (95% CI 0.034 to 0.065; I^2=57%; Figure 7).[20,36,48,56,60,61] Results were similar when the analysis was performed using the profile likelihood method. The mean difference in AUROC was slightly higher in the one study of children (0.121, 95% CI 0.068 to 0.174)[56,61] than in four studies of mixed populations of adults and children (0.046, 95% CI 0.038 to 0.054, I^2=0%),[20,36,48,60] but there was no statistically significant interaction with age group (p=0.07). Differences in how severe brain injury was defined could explain some of the differences in estimates. The study in children used a composite outcome of head CT imaging findings or need for intervention.[61] All of the studies of mixed populations of adults and children defined severe brain injury on the basis of CT imaging findings.

Figure 7. Pooled AUROC of severe brain injury for the total Glasgow Coma Scale versus the motor component only



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; mGCS=motor Glasgow Coma Scale; n=number; NR=not reported; tGCS=total Glasgow Coma Scale

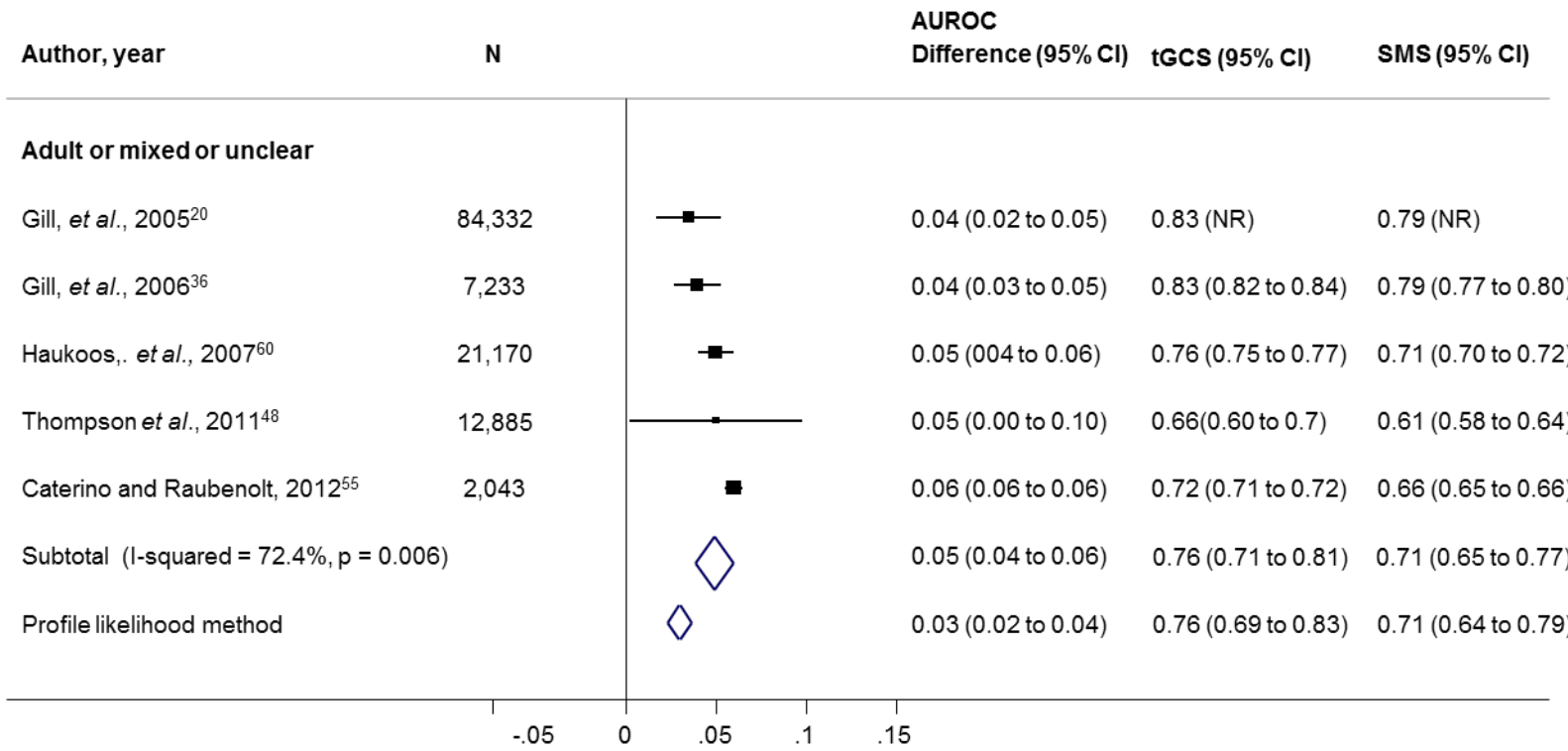
Results were similar when studies were stratified according to whether they used out-of-hospital or ED GCS scores, or when analyses were restricted to studies conducted in the United States, studies that focused on TBI patients, or low risk of bias studies (Table 4). There were also no differences when we excluded studies with potential overlap.[20,60] No study was conducted on the NTDB database.

One study found no difference between out-of-hospital tGCS (cutoff of ≤13) versus the mGCS (cutoff of ≤5) in sensitivity (62%, 95% CI 55 to 68 vs. 61%, 95% CI 54 to 67) or specificity (85%, 95% CI 83 to 88 vs. 89%, 95% CI 88 to 91) for identifying patients with severe head injury (defined as head AIS score of ≥4)[19] (Table 5).

Total Glasgow Coma Scale Versus Simplified Motor Scale

The tGCS was slightly better than the SMS at discriminating patients found to have severe brain injury from those without a severe brain injury; results were very similar to the comparison of tGCS versus mGCS. Based on five studies, the pooled AUROC for the tGCS was 0.763 (95% CI 0.710 to 0.815) and for the mGCS was 0.713 (95% CI 0.654 to 0.771), with a mean difference of 0.048 (95% CI 0.038 to 0.059, I^2=72%; Figure 8).[20,36,48,55,60] Although statistical heterogeneity was present, the estimates from individual studies were similar (mean difference in AUROC ranged from 0.035 to 0.060), and all studies favored the tGCS. All of the studies defined severe brain injury similarly, based on head CT imaging findings. All of the studies were conducted in the United States and none focused on children or patients with TBI. There were no differences in estimates between studies that utilized out-of-hospital GCS scores or ED GCS scores (Table 4). Results were also unchanged when analyses were restricted to low risk of bias studies, or when studies with potential overlap[20,60] were excluded. No study utilized the NTDB database. Results were unchanged using the profile likelihood method.

Figure 8. Pooled AUROC of severe brain injury for the total Glasgow Coma Scale versus the Simplified Motor Scale



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; n=number; NR=not reported; SMS=Simplified Motor Scale; tGCS=total Glasgow Coma Scale

One study found out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for severe brain injury based on presence of head CT imaging findings (45%, 95% CI 44 to 46 vs. 41%, 95% CI 40 to 42) and similar specificity (89%, 95% CI 89 to 90 vs. 90%, 95% CI 90 to 91; Table 5).[55]

Motor Glasgow Coma Scale Versus Simplified Motor Scale

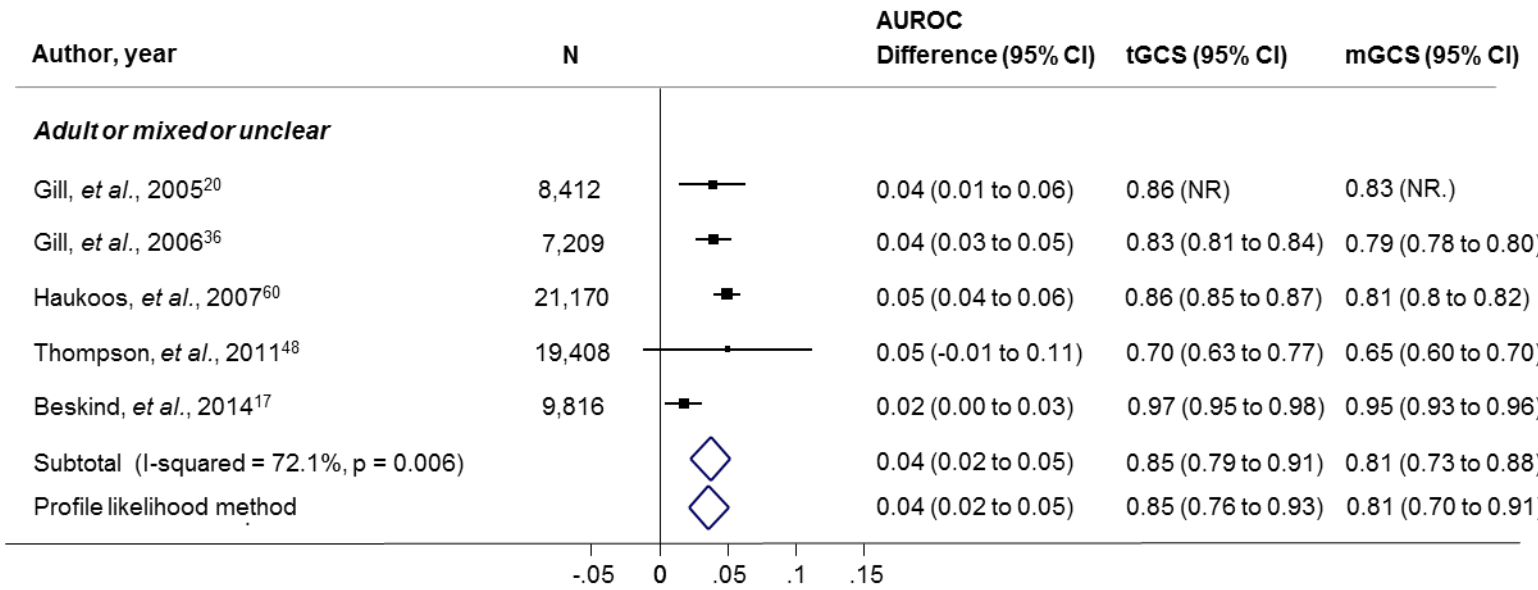
There was no difference between the mGCS versus the SMS in ability to discriminate patients who would undergo a neurosurgical intervention from those who would not undergo a neurosurgical intervention (4 studies, mean difference in AUROC 0.000, 95% CI -0.008 to 0.007, I^2=0%).[20,36,48,60] There was no statistical heterogeneity and findings were unchanged in sensitivity and stratified analyses.

Emergency Intubation

Total Glasgow Coma Scale Versus Motor Glasgow Coma Scale

The tGCS was slightly better than the mGCS at discriminating patients who underwent emergency intubation from those who did not undergo intubation. Based on five studies, the pooled AUROC for the tGCS was 0.851 (95% CI 0.794 to 0.908) and for the mGCS was 0.807 (95% CI 0.735 to 0.880), with a mean difference of 0.038 (95% CI 0.023 to 0.052; I^2=72%; Figure 9).[17,20,36,48,60] Although statistical heterogeneity was present, estimates were similar across studies (mean difference in AUROC ranged from 0.018 to 0.050) and all studies favored the tGCS. There were no clear difference in estimates between two studies that focused on intubation out-of-hospital (mean difference in AUROC 0.039 and 0.040)[20,36] and three studies that evaluated any emergency intubation (ED or out-of-hospital) (mean difference in AUROC 0.018 to 0.050).[17,48,60] All of the studies evaluated mixed populations of adults and children and were conducted in the United States. There were no differences when studies were stratified according to use of out-of-hospital or ED GCS scores, when analyses were restricted to low risk of bias studies, or when we excluded studies with potential overlap in populations. One study reported subgroup findings for trauma patients with TBI;[17] as in the analysis of patients with any trauma, results favored the tGCS (mean difference in AUROC 0.011, 95% CI -0.010 to 0.032). No study was based on data from the NTDB.

Figure 9. Pooled AUROC of intubation for the total Glasgow Coma Scale versus the motor component only

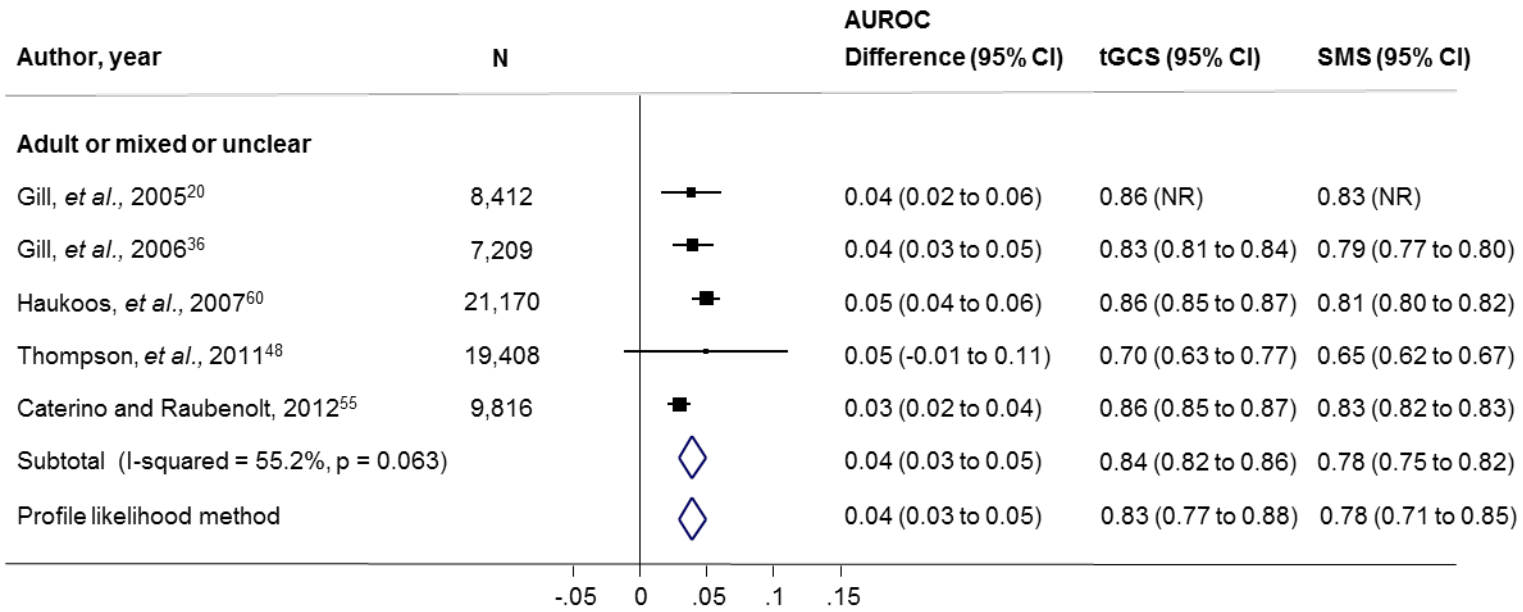


AUROC=area under the receiver operating characteristics curve; CI=confidence interval; mGCS=motor Glasgow Coma Scale; n=number; NR=not reported; tGCS=total Glasgow Coma Scale

Total Glasgow Coma Scale Versus Simplified Motor Scale

The tGCS was slightly better than the SMS at discriminating patients who underwent emergency intubation from patients who did not undergo intubation; results were very similar to the comparison of tGCS versus mGCS. Based on five studies, the pooled AUROC for the tGCS was 0.843 (95% CI 0.823 to 0.864) and for the SMS was 0.783 (95% CI 0.747 to 0.819), with a mean difference of 0.040 (95% CI 0.030 to 0.050, I^2=55%, Figure 10).[20,36,48,55,60] Although statistical heterogeneity was present, the estimates from individual studies were similar (mean difference in AUROC ranged from 0.030 to 0.050), and all studies favored the tGCS. All of the studies defined severe brain injury similarly, based on head CT imaging findings. All of the studies were conducted in the United States and none focused on children or patients with TBI. There were no differences in estimates between studies that utilized out-of-hospital GCS scores or ED GCS scores (Table 4). Results were also unchanged when analyses were restricted to low risk of bias studies, or when studies with potential overlap[20,60] were excluded. No study utilized data from the NTDB. Results were unchanged using the profile likelihood method.

Figure 10. Pooled AUROC of intubation for the total Glasgow Coma Scale versus the Simplified Motor Scale



AUROC=area under the receiver operating characteristics curve; CI=confidence interval; n=number; NR=not reported; SMS=Simplified Motor Scale; tGCS=total Glasgow Coma Scale

One study found the out-of-hospital tGCS (cutoff of ≤13) associated with slightly higher sensitivity than the SMS (cutoff of ≤1) for identifying patients who underwent emergency intubation (76%, 95% CI 74 to 77 vs. 73%, 95% CI 71 to 74) and slightly lower specificity (89%, 95% CI 89 to 89 vs. 91%, 95% CI 90 to 91; Table 5).[55]

Motor Glasgow Coma Scale Versus Simplified Motor Scale

There was no clear difference between the mGCS versus the SMS in ability to discriminate between patients who undergo emergency intubation from those who do not undergo emergency intubation (4 studies, mean difference in AUROC 0.000, 95% CI -0.007 to 0.007, I^2=0%).[20,36,48,60] There was no statistical heterogeneity and findings were unchanged in sensitivity and stratified analyses.

Trauma Center Need

One study that utilized NTDB data (n=811,143) evaluated the predictive utility of the tGCS versus the mGCS for identifying patients in need of trauma center care (defined as ISS >5, ICU admission >24 hours, need for urgent surgery, or death in the ED).[54] Differences in the AUROC (0.62 vs. 0.61), sensitivity (30% vs. 27%), and specificity (93% vs. 95%) were small (statistical significance not reported). The adjusted risk estimates were also similar (odds ratio [OR] 3.03, 95% CI 2.95 to 3.13 for tGCS score of ≤13 vs. >13 and OR 3.37, 95% CI 3.27 to 3.48 for mGCS score of ≤5 vs. >5).

Other Outcomes

One study (n=104,035) of children with TBI in the NTDB found the tGCS was better able to discriminate those with major injury (defined as an ISS of >15) from those without major injury (AUROC 0.720, 95% CI 0.715 to 0.724 vs 0.681, 95% CI 0.677 to 0.686).[56] Another study (n=2231) of children with TBI found that the tGCS was better able to discriminate those admitted to the ICU from those not admitted to the ICU (AUROC 0.772, 95% CI 0.754 to 0.790 vs. 0.721, 95% CI 0.705 to 0.738, p<0.001), those with a length of stay 5 days or longer from those with a stay of more than 5 days (0.683, 95% CI 0.660 to 0.706 vs. 0.644, 95% CI 0.622 to 0.666, p<0.001)), those discharged to rehabilitation from those not discharged to rehabilitation (0.804, 95% CI 0.782 to 0.826 vs. 0.766, 95% CI 0.740 to 0.792, p<0.001)), and those dependent on a caregiver from those not dependent on a caregiver following discharge (0.757, 95% CI 0.732 to 0.783 vs. 0.747, 95% CI 0.722 to 0.772, p=0.06).[52]

Key Question 1a. How does predictive utility vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medications in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Points

- Age: Differences in the AUROC between the tGCS versus the mGCS were similar in studies that enrolled children and those that enrolled mixed populations of adults and children (SOE: Low).
- Type of trauma: Differences in the AUROC between the tGCS versus the mGCS were similar in studies that evaluated patients with TBI and those that enrolled mixed trauma patients (SOE: Low).
- Out-of-hospital vs. ED assessment: One study of adults found no differences between out-of-hospital and ED GCS scores on discrimination for mortality or neurosurgical intervention but another study of adults or children found out-of-hospital GCS scores associated with higher discrimination for mortality than ED scores (AUROC 0.754 vs. 0.635, p not reported). Differences in the AUROC between the tGCS versus the mGCS were similar in studies that evaluated out-of-hospital GCS scores and those that used ED scores (SOE: Insufficient).
- No study evaluated how intoxication status, blood pressure, intubation status, receipt of field intubation, or level/training of field assessors impacts comparative predictive utility of the tGCS versus the mGCS or SMS.

Detailed Assessment

Age

No study that evaluated mixed populations of adults and children performed analyses stratified according to age group. Among the head-to-head studies on predictive utility, two studies on mortality, one study on neurosurgical interventions, and two studies on severe brain injuries focused on children. For all of these outcomes, differences between the tGCS versus the mGCS in the AUROC slightly favored the tGCS and estimates were similar in studies that evaluated children and studies that evaluated mixed populations (Table 4). For mortality, the mean difference in AUROC was 0.006 (95% CI 0.002 to 0.011, I^2=0%) in two studies of children[52,56] and 0.017 (95% CI 0.015 to 0.020, I^2=0%) in nine studies of adults.[17,18,20,36,48,53,58-60] For neurosurgical intervention, results were similar in one study[52] of children (mean difference in AUROC 0.034, 95% CI 0.009 to 0.059) and five studies of mixed populations of adults and children (mean difference in AUROC 0.026, 95% CI 0.019 to 0.034, I^2=0%). For severe brain injury, the mean difference in AUROC was higher in one study of children (0.121, 95% CI 0.068 to 0.174)[61] than in four studies of mixed populations of adults and children (0.046, 95% CI 0.038 to 0.054, I^2=0%),[20,36,48,60] but there was no statistically significant interaction with age (p=0.07). The study in children used a broader definition for severe brain injury differently (based on head

CT imaging findings or need for intervention)[61] than the studies conducted in mixed populations, which focused on CT imaging findings. One study of children up to 18 years of age found no clear differences in AUROC estimates for mGCS between the subgroup of children 0 to 3 years of age and the entire cohort for survival to hospital discharge (0.936, 95% CI 0.911 to 0.962 vs. 0.941, 95% CI 0.926 to 0.957), craniotomy (0.659, 95% CI 0.597 to 0.721 vs. 0.638, 95% CI 0.601 to 0.675), or ICU admission (0.723, 95% CI 0.696 to 0.750 vs. 0.721, 95% CI 0.705 to 0.738), or length of stay greater than 4 days (0.589, 95% CI 0.555 to 0.623 vs. 0.644, 95% CI 0.622 to 0.666). The mGCS was associated with slighter better discrimination for being dependent on caregiver in those 0 to 3 years (0.787, 95% CI 0.752 to 0.821 vs. 0.747, 95% CI 0.722 to 0.772).[52]

Two studies of children with TBI found a tGCS score less than 13 or 13 or less associated with sensitivity of 80 percent and 100 percent for mortality; specificity was 74 percent[64] and 86 percent (Appendix I).[65] In three studies of adults or mixed populations of adults and children, sensitivity of the tGCS (cutoff of ≤13) ranged from 71 percent to 80 percent and specificity from 68 percent to 88 percent (Table 5).[19,53,55]

One study evaluated the diagnostic accuracy of tGCS in older (≥70 years) versus younger (<70 years) adults (Appendix I).[66] Based on a tGCS cutoff of 13 or less, it found that sensitivity of the tGCS was lower in adults 70 years or older and worse by more than 20 percent versus those younger for mortality (51% vs. 86%), severe TBI (28% vs. 53%), neurosurgical intervention (43% vs. 66%), and emergency intubation (58% vs. 78%). Specificity was greater than 90 percent on these outcomes in adults 70 years or older and 5 percent to 10 percent higher than adults younger than 70 years. At a cutoff of 14 or less, sensitivity improved in patients 70 years or older by about 10 percent on all outcomes and specificity decreased by 5 percent to 10 percent, but the differences versus patients younger than70 years of age remained similar. In older adults, decreases in the tGCS to 14 from 15 (adjusted OR [AOR] 1.40, 95% CI 1.07 to 1.83) and from 13 to 14 (AOR 2.34, 95% CI 1.57 to 3.52) were associated with greater risk of mortality than in adults younger than 70 years of age (AOR 1.22, 95% CI 0.88 to 1.71 and 1.45, 95% CI 0.91 to 2.30, respectively).

Type of Trauma

No study that evaluated mixed populations of trauma patients performed analyses stratified according to the type of trauma. Among the head-to-head studies on predictive utility, differences in the AUROC between the tGCS versus the mGCS were similar among studies that focused on patients with TBI versus those that evaluated patients with mixed trauma (Table 4). For mortality, the difference in the AUROC was 0.009 (95% CI -0.002 to 0.020) in three studies of TBI patients[17,52,58] and 0.014 (95% CI 0.007 to 0.020) in nine studies of mixed trauma patients.[17,18,20,36,48,53,56,59,60] For neurosurgical intervention, the difference in the AUROC was 0.017 (95% CI -0.022 to 0.056) in two studies of TBI patients[17,52] and 0.026 (95% CI 0.019 to 0.034) in five studies of mixed trauma patients.[17,20,36,48,60] For emergency intubation, the difference in the AUROC was 0.011 (95% CI -0.010 to 0.032) in one study of TBI patients[17] and 0.038 (95% CI 0.023 to 0.052) in five studies of mixed trauma patients.[17,20,36,48,60]

No head-to-head study evaluated the predictive utility of the tGCS versus the SMS specifically in patients with TBI, or evaluated effects of type of TBI or presence of other injuries on predictive utility. One study of patients with multiple injuries (based on head and skeletal injury AIS scores ≥3) found the tGCS (cutoff of ≤12) associated with sensitivity of 87 percent (95% CI 78 to 94) and specificity of 71 percent (95% CI 68 to 74, Appendix I).[67] One study of

TBI patients found the tGCS associated with an AUROC for mortality for 0.85 (95% CI 0.80 to 0.90) among all patients and 0.82 (95% CI 0.77 to 0.86) among the subgroup of patients with a GCS score less than 15 (Appendix I).[68]

Field Versus Emergency Department Assessment

Among the head-to-head studies on predictive utility, differences between the tGCS versus the mGCS or the tGCS versus the SMS were similar among studies that used out-of-hospital GCC scores versus ED GCS score (Table 4). For mortality, the difference in AUROC for studies that used out-of-hospital GCS scores was 0.012 (95% CI 0.0004 to 0.020, I^2=74%) and for studies that used ED GCS scores was 0.020 (95% CI 0.006 to 0.034). Findings were similar for other outcomes and for comparisons of tGCS versus SMS. Of the studies that were pooled, two that used out-of-hospital GCS scores[36,48] and two that used ED scores[20,60] were conducted in the same trauma center/system. In these studies, differences in the AUROC between the tGCS versus the mGCS or SMS were very similar when results based on out-of-hospital versus ED GCS scores were compared from each system. For example, for mortality, a study from the Loma Linda trauma center found out-of-hospital tGCS associated with an AUROC for mortality of 0.89 for the tGCS versus 0.88 for the mGCS and 0.86 for the SMS.[36] Using ED scores, the AUROCs were 0.906, 0.894, and 0.878, respectively.[20] A study from the Denver trauma system found AUROCs for mortality of 0.90, 0.88, and 0.87 for the tGCS, mGCS, and SMS (missing GCS data excluded from analysis).[48] AUROCs were similar in a study from the Denver trauma system that used ED GCS scores (0.92, 0.90, and 0.89, respectively).[60] The degree to which the patient populations in studies conducted in the same trauma system overlapped in the studies was unclear . Sample sizes were not the same in the out-of-hospital and ED studies from the same trauma center/system, indicating that some patients with out-of-hospital GCS scores did not have had ED scores, or vice versa. In addition, in the Denver studies, data collection dates were not identical for the out-of-hospital (1999 to 2008)[48] and ED (1995 to 2004)[60] studies.

Three studies directly compared discrimination for out-of-hospital versus ED GCS scores obtained from the same patients (Appendixes H and I). One study of adults found no differences between out-of-hospital versus ED tGCS for mortality (AUROC 0.84 vs. 0.84) or for neurosurgical intervention (0.80 vs. 0.83).[58] One study (adults or children) found field mGCS scores associated with better discrimination for mortality than ED mGCS scores (0.754 vs. 0.635, p not reported).[69] One other study of children found no difference between out-of-hospital versus ED tGCS scores (cutoff of <13) in sensitivity (80% vs. 80%) or specificity (74% vs. 76%) for mortality.[64]

Other Factors

No study evaluated how intoxication status, blood pressure, intubation status, receipt of field intubation, or level/training of field assessors impacts comparative predictive utility of the tGCS versus the mGCS or SMS.

Key Question 2. In patients with known or suspected trauma, what are the comparative effects of the tGCS compared with the mGCS on over- and under-triage (e.g., proportion of patients misclassified with regard to measures of injury severity or need for early interventions and transport to a lower versus higher level of care)?

Key Points

- No study compared effects on the tGCS versus the mGCS or SMS on rates of over- or under-triage.

Detailed Assessment

No study compared effects of the tGCS versus the mGCS or SMS on rates of over- or under-triage. The head-to-head studies in Key Question 1 of tGCS versus the mGCS or SMS were not designed to assess effects on over- or under-triage, because all patients underwent a single triage decision based on the tGCS, and do not reflect differences in actual triage decisions. In additions, such studies do not account for other factors that impact triage decisions.[4]

Nonetheless, measures of diagnostic accuracy may provide some indirect information about the potential degree of over-triage (1-specificity, or the false-positive rate) and under-triage (1-sensitivity, or the false-negative rate). However, data on diagnostic accuracy were very limited. For mortality, three studies found that differences in sensitivity between the tGCS and mGCS ranged from 0 percent to 1 percent and differences in specificity ranged from 0 percent to 5 percent (Table 5). For other outcomes, data on sensitivity and specificity differences were limited to a single study.

One study on predictive utility included in Key Question 1 attempted to estimate the net effect on over- and under-triage.[54] It found that replacing the tGCS (cutoff of ≤13) with the mGCS (cutoff of ≤5) in the National Trauma Triage Protocol (NTTP) would result in a net decrease in over- or under-triage of 0.4 percent (prevent over-triage in 1.7 percent and under-triage in 0.2 percent, while causing over-triage in 0.4 percent and under-triage in 1.1 percent), based on accuracy for identification of patients with trauma center need (ISS of >15, ICU admission ≥24 hours, need for urgent surgery, or death in the ED). In addition, it found that benefits of the mGCS on rates of over- or under-triage may be higher, as 0.5 percent of patients who would have been under-triaged by mGCS would have received the appropriate triage decision based on other elements of the NTTP. Another study included in Key Question 1 found that based on diagnostic accuracy, rates of misclassification for in-hospital survival were very similar for the tGCS versus the mGCS (4.9% vs. 5.1%).[18]

Key Question 2a. How do effects on clinical decisionmaking vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Points

- No evidence.

Key Question 3. In patients with known or suspected trauma, what is the comparative effectiveness of the tGCS compared with the mGCS on clinical outcomes (e.g., mortality, morbidity, quality of life)?

Key Points

- No study compared effects on the tGCS versus the mGCS or SMS on clinical outcomes.

Key Question 3a. How do effects on clinical outcomes vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Points

- No evidence.

Key Question 4. In patients with known or suspected trauma, what is the comparative reliability (e.g., interrater and intra-rater kappa) and ease of use (e.g., time to complete, amount of missing data, user reported satisfaction) of the tGCS compared with the mGCS score?

Key Points

- The interrater reliability of tGCS and mGCS appears to be high, but evidence was insufficient to determine if there were differences between scales (SOE: Insufficient).
- Three studies found the tGCS associated with a lower proportion of correct scores than the mGCS (differences in proportion of correct scores ranged from 6% to 27%), though the difference was statistically significant in only one study (SOE: Low).

- Three studies found that training or use of a scoring aid increased the proportion of correct scores on both the tGCS and mGCS (increase in proportion of correct scores ranged from 32% to 70%) (SOE: Low).

Detailed Assessment

Interrater Reliability

Interrater or inter-observer reliability refers to the extent to which different people using a scale arrive at the same rating for the same patient. Two studies evaluated the interrater reliability of the GCS in TBI patients presenting to the ED (Table 6, Appendix J).[61,70]

Table 6. Reliability and ease of use findings for comparisons of tGCS and mGCS

	Author, year Study Design Risk of Bias	Objective	N Type of Raters	Results Overall	Results by Patient Characteristics	Results by Provider Characteristics
Interrater Reliability	Holmes, *et al.,* 2005[61] Cross-sectional High	Compare the pediatric GCS in children ≤2 years to the standard GCS in children >2 years.	N=102 patients Emergency Physicians	**Weighted kappa (95% CI) across raters** tGCS: 0.77 (0.38 to 1.00) for ≤2 year olds and 0.91 (0.75 to 1.00) for >2 year olds mGCS: 0.91 (0.75 to 1.00) for ages combined	**Weighted kappa (95% CI) across raters** tGCS: 0.77 (0.38 to 1.00) for ≤2 year olds 0.91 (0.75 to 1.00) for >2 year olds mGCS: not reported by age groups	None reported
Ease of Use/Correct Scoring	Bledsoe, *et al.,* 2015[71] Cross-sectional Low	Evaluate tGCS and its components in standardized video vignettes.	N= 217 providers; 10 patient scenarios AEMT CCP EMT Nurses Paramedics Physicians Residents	**Correct scores (95% CI), tGCS vs. mGCS** Across all vignettes and participants: 33.1% (30.2 to 36.0) vs. 59.8% (58.1 to 61.5)	**Correct Scores** **tGCS** Accuracy was lowest for scenarios with tGCS scores of 9 to 13 (<20%; data taken from figure) **mGCS** Accuracy not reported by score or severity	**Correct scores (95% CI) tGCS vs. mGCS by provider type** Highest/best residents: 51% (44.5 to 57.5) vs. 78% (71 to 84.5) Lowest/worst nurses tGCS: 29% (10.3 to 47.7) EMTs mGCS: 51% (43.7 to 58.3)
Ease of Use/Correct Scoring	Feldman, *et al.,* 2015[72] RCT Low	Assess ability of EMS personal to correctly score the tGCS and its components and to determine if scoring improves with the use of a scoring aid.	N=178 Providers EMTs Paramedics	**Correct scores, tGCS vs. mGCS** All scenarios: 41.0% vs. 50.6%	**Correct, tGCS vs. mGCS** Mild TBI scenarios: 54.2% vs. 74.6% Moderate TBI scenarios: 28.8% vs. 35.6% Severe TBI scenarios: 40.0% vs. 41.7%	None reported

	Author, year Study Design Risk of Bias	Objective	N Type of Raters	Results Overall	Results by Patient Characteristics	Results by Provider Characteristics
Ease of Use/Correct Scoring	Heim, et al., 2009[73] Cross-sectional High	Assess knowledge of GCS and scoring of a clinical scenario.	N=103 providers; 1 patient scenario Air rescue physicians	**Incorrect (correct) scores** tGCS: 36.9% (63.1%) mGCS: 27.2% (72.8%)	None reported	**Incorrect (correct) scores by experience** Registrars (trainees): 47.5% (52.5%) p=0.095 Fellow: 33.3% (66.7%) p=0.671 Consultant: 0% (100%) p<0.05 Private practice: 36.8% (63.2%) p=1.00 Specialty was not associated with statistically significant differences in errors (anesthesia, internal medicine, general practice, other)
Field vs. ED Agreement	Kerby, et al., 2007[74] Cross-sectional High	Linkage of EMS and trauma registry data to determine if differences may cause inappropriate enrollment in out-of-hospital trials.	N=3,052 patients EMTs all levels ED personnel not specified	**Weighed kappa (95% CI) tGCS vs. mGCS:** 0.53 (0.48 to 0.58) vs. 0.48 (0.43 to 0.53)	**Weighted kappa (95% CI) by Transport time tGCS vs. mGCS** <20 minutes: 0.56 (0.50 to 0.61) vs. 0.52 (0.46 to 0.57) ≥20 minutes: 0.42 (0.32 to 0.52) vs. 0.35 (0.25 to 0.46)	None reported

CCP=critical care paramedic; CI=confidence interval; ED=emergency department; AEMT=advanced emergency medical technician; EMS=emergency medical services; EMT=emergency medical technician; GCS=Glasgow Coma Scale; KQ=Key Question; mGCS=motor Glasgow Coma Scale; n=number; RCT=randomized controlled trial; TBI=traumatic brain injury; tGCS=total Glasgow Coma Scale; vs=versus

One high risk of bias study (n=102 patients; number of raters unclear) conducted in the United States compared the interrater reliability of a pediatric version of the tGCS in preverbal children 2 years old and younger and the standard version in children older than two.[61] Two faculty emergency physicians assessed the same patients upon presentation to the ED. The agreement on tGCS was higher for older children using the standard tGCS (weighted kappa 0.91, 95% CI 0.75 to 1.00) than for younger children using the pediatric tGCS (weighted kappa 0.77, 95% CI 0.38 to 1.00), but estimates were imprecise, particularly for younger children, and CIs overlapped. The interrater reliability of the mGCS for all children in the sample (not reported by age group) was high (weighted kappa 0.91, 95% CI 0.75 to 1.00).

A moderate risk of bias study conducted in Japan assessed interrater reliability of the tGCS in 66 patients with suspected TBI.[70] The tGCS was assessed by two members of the medical team (number of assessors=33) upon arrival at the ED. The weighted kappa for the TBI patients was 0.74 (95% CI 0.71 to 0.76). The weighted kappa for all trauma patients was not reported.

Methodological shortcomings in the studies include lack of information on how patients were selected or use of a convenience sample, unclear blinding of assessors to other assessors' ratings, and unclear timing of GCS assessment.

Ease of Use

Five studies (reported in 4 articles) evaluated the ability of medical personal to correctly score the tGCS and the mGCS,[71-73,75] based on simulated patients presented in video[71,75] or written[72,73] scenarios in which correct scores were determined by experts (Appendix J). In three studies, the assessors were Emergency Medical Technicians [24] and paramedics or paramedic students,[72,75] in one study the assessors were air rescue physicians,[73] and the fourth study evaluated assessors with various types and levels (e.g., Emergency Medical Technician [EMT], nurse, physician, and resident) of training.[71] Four studies were conducted in the United States[71,72,75] and one in Switzerland.[73] Sample sizes ranged from 46 to 217 providers/raters. In four studies, the scenarios represented a spectrum of injury severity from mild to severe; the fifth study[73] used a single scenario in which the correct GCS score was 6. Three studies[71-73] evaluated the proportion of correct scores with the tGCS and the mGCS and the other two[75] only evaluated the tGCS.

The studies differed in how they were designed. One study was a randomized controlled trial (RCT) (rated low risk of bias) that compared tGCS and mGCS scores with versus without the use of a scoring aid.[72] One article reported two studies (both rated moderate risk of bias): the first was a before-after study on the effects of video training on tGCS scoring, and the second was a before-after study on the effects of video training on tGCS scoring, in which participants were also randomized to use of a GCS scoring aid.[75] The other two studies (one rated low risk of bias[71] and the other high risk of bias)[73] used a cross-sectional design. The RCT and one cross-sectional study were rated low risk of bias.[71,72] Methodological shortcomings in the high and moderate risk of bias studies included use of the same patient scenarios on repeat testing and unclear methods for determining correct answers;[75] the high risk of bias study only evaluated one scenario and allowed assessors 10 minutes to rate a written scenario.[73]

Three studies consistently found that the proportion of correct scores was lower with the tGCS than the mGCS.[71-73] One cross-sectional study found that the overall proportion of correct scores by assessors from different types and levels of training was 33.1 percent (95% CI 30.2 to 36.0) with the tGCS and 59.8 percent (95% CI 58.1 to 61.5) with the mGCS.[71] In two other studies, the proportion of correct scores was lower with the tGCS than the mGCS, but

differences were not statistically significant. A cross-sectional study found that the proportion of correct scores by air rescue physicians was 63.1 percent (53.8% to 72.4%) for the tGCS and 72.8 percent (95% CI 64.2% to 81.4%) for the mGCS.[73] An RCT found that the proportion of correct scores by EMTs or paramedics was lower using the tGCS than the mGCS in assessors randomized to a scoring aid (56.7%, 95% CI 46.5 to 66.9 vs. 70.0%, 95% CI 60.5 to 79.5) as well as those randomized to no aid (25.0%, 95% CI 16.0 to 34.0 vs. 30.7%, 95% CI 21.1 to 40.3), though differences were not statistically significant.[72] In this study, the overall rate of correct tGCS scores was 41 percent and 69 percent of scores were within 1 point of the correct score. No other study reported the degree to which incorrect scores differed from correct scores (e.g., the proportion of scores that were incorrect by 1 point vs. those that differed by ≥2 points), or the proportion of scores that cross GCS triage thresholds (e.g., patient scenario in which the correct tGCS score is 14 or 15 is scored as ≤13).

Three studies (reported in 2 articles) evaluated the effect of training and scoring aids on performance of the tGCS and the mGCS.[72,75] In one study, 75 attendees at an Emergency Medical Services meeting evaluated four video scenarios depicting a spectrum of injury severity before and after watching a 13-minute training video.[75] Across the patient scenarios, the proportion of correct scores on the tGCS increased from 14.7 percent before the video to 64 percent after the video ($p<0.001$); a similar pattern was observed for each individual patient scenario. In a second study reported in the same article, 46 students in a paramedic class watched the same video, and in addition were randomized to use or not use a tGCS reference card (showing the scoring for each GCS component). The proportion of correct tGCS scores improved both among those randomized to the reference card aid (50% pre-video and 100% post-video, $p=0.001$) as well as those randomized to no reference card aid (7.7% pre-video and 76.9% post-video, $p<0.0001$). In the third study, EMTs and paramedics (n=178) scored one out of nine possible written scenarios (depicting TBI and a spectrum of injury severity) after randomization to a scoring aid or no scoring aid.[72] The proportion of correct scores was higher with the aid versus without the aid for both the tGCS (56.7% vs. 25.0%, difference 31.9%, 95% CI 18.3 to 45.6) and the mGCS (70.0% vs. 30.7%, difference 39.7%, 95% CI 26.2 to 53.1).

No study compared the ease of use of the tGCS or the mGCS in terms of the amount of time needed to complete assessments of trauma patients, or satisfaction of assessors.

Key Question 4a. How do comparative reliability and ease of use vary according to patient age or other patient characteristics (e.g., TBI vs. unspecified or other trauma, systolic blood pressure, level of intoxication, type of trauma, intubation or receipt of medication in the field), the training and background of the person administering the instrument, and the timing/setting of assessment (i.e., in the field vs. upon presentation to the ED or urban vs. rural location)?

Key Points

- Evidence was insufficient to assess effects of patient or assessor characteristics on comparative interrater reliability of the tGCS versus the mGCS (SOE: Insufficient).
- The proportion of correct GCS scores was generally lowest for assessment of patient scenarios with moderate injury severity in three studies, including one study that evaluated the tGCS and the mGCS (SOE: Low).

- Evidence was insufficient to determine effects of level of training or professional background on the proportion of correct scores on the tGCS versus the mGCS (SOE: Insufficient).
- No study evaluated how comparative interrater reliability or ease of use of the tGCS versus the mGCS vary according to assessment setting (SOE: Insufficient).
- One study found agreement between out-of-hospital and ED scores was similar for the tGCS and the mGCS (SOE: Low).

Detailed Assessment

Ease of Use

Patient Characteristics

Evidence regarding how patient or assessor characteristics impact ease of use of GCS scales was limited. Three studies that assessed the proportion of correct scores on the tGCS or mGCS varied according to TBI severity in the patient scenario assessed.[71,72,75] In one study (75 assessors) the proportion of correct scores were 64.0 percent to 76.0 percent before viewing a training video and 89.3 percent to 98.6 percent using two patient scenarios of tGCS scores of 15.[75] For the scenarios with a tGCS of 5, the proportion correct increased from 45.3 percent pre-video to 94.7 percent post-video, and in the scenario with a tGCS of 8, the proportion correct increased from 36.0 percent to 74.7 percent. In another study (178 assessors), the percent correct for both the tGCS and mGCS was highest without a scoring aid in the three mild TBI scenarios (proportion correct 44.8% for the tGCS and 58.6% for the mGCS) and lower in the three severe TBI scenarios (20% and 13.3%, respectively) and the three moderate TBI scenarios (10.3% and 20.7%, respectively).[72] With a scoring aid, the proportion of correct scores at all severity levels improved but still remained highest for mild TBI (63.3% for tGCS and 90.0% for mGCS) and lower in the moderate (46.7% and 50.0%, respectively) and severe scenarios (60.0% and 70.0%, respectively). The third study (217 assessors) found that the proportion of correct scores was lowest for scenarios in which the tGCS scores were 9 to 13 (<20%) and highest for scenarios in which the tGCS score was 3 (>90% correct).[71]

Rater Characteristics

Evidence on how assessor characteristics such as type or level or training impacts the proportion of correct GCS scores was limited. One study (46 assessors) found that prior to watching a training video, neither level of training nor participation in a training course was associated with correct scoring.[75] After watching the video, participants who had taken a trauma care course in the last 5 years were more likely to score the scenarios correctly (p=0.001). Another study evaluated the proportion of correct scores according to training background (advanced EMTs, EMTs, critical care paramedics, paramedics, nurses, resident physicians, and staff physicians).[71] The proportion of correct scores was highest for residents with both the tGCS (51%, 95% CI 44.5 to 57.5) and the mGCS (78%, 95% CI 71.5 to 84.5) than the other assessor categories. The proportion of correct scores on the tGCS was lowest for nurses (29%, 95% CI 10.3 to 47.7) and the proportion of correct mGCS scores was lowest for EMTs (51%, 95% CI 43.7 to 58.3). A third study found that the highest rate of errors in scoring on the tGCS was in registrars (physicians who had not yet obtained the rank of consultant) at 47.5 percent, compared with 33.3 percent in fellows and 36.8 percent of physicians in private practice.[73] Consultants had

no errors, but the sample size was small (n=8). There were no clear differences in the rate of errors between different specialties (anesthesia, internal medicine, general practice, or others).

Field Versus Emergency Department Agreement

No study evaluated how interrater reliability or ease of use vary according to assessment setting. Four studies evaluated agreement between out-of-hospital and ED GCS scores.[64,74,76,77] One high risk of bias study (n=3,052) found that agreement between out-of-hospital and ED scores for adults with blunt or penetrating trauma were similar for the tGCS (weighted kappa 0.53, 95% CI 0.48 to 0.58) and the mGCS (weighted kappa 0.48, 95% CI 0.43 to 0.53).[74] Agreement between out-of-hospital and ED scores was somewhat higher among patients with shorter versus those with longer transport time (weighted kappa for tGCS 0.56 vs. 0.42, respectively; weighted kappa for mGCS 0.52 vs. 0.35, respectively).

The three other studies focused on the tGCS without a comparison with the mGCS and generally found high levels of agreement between out-of-hospital and ED scores.[64,76,77] One high risk of bias study (n=7,823) found no statistically significant differences between out-of-hospital and ED tGCS scores for adult trauma patients, with 82 percent of tGCS scores falling in the same category (mild 14-15, moderate 9-13, or severe 3-8), 3 percent higher GCS (less severe in ED) compared with out-of-hospital, and 15 percent lower GCS (more severe in ED).[76] A moderate risk of bias study (n=1,181) found good agreement between out-of-hospital and ED tGCS scores for trauma patients 15 years old or older, based on an intra class correlation coefficient of 0.74 (95% CI, 0.37 to 1.12).[77] In this study, 96.3 percent of out-of-hospital-ED pairs were within a predetermined range of acceptability of 3 points. Another moderate risk of bias study (n=185) found good agreement between out-of-hospital and ED tGCS scores in children, with a weighted kappa of 0.74 (95% CI 0.63 to 0.85).[64] Although ED scores tended to be higher than out-of-hospital scores, differences were small (0.44 points on average, with no difference in median scores).

Discussion

Key Findings and Strength of Evidence

The Key Findings of this review, with overall strength of evidence ratings, are summarized in Appendix G. Details about ratings for individual strength of evidence domains are shown in Appendix F.

Based on head-to-head studies, we found that the total Glasgow Coma Scale (tGCS) is associated with slightly better predictive utility than the motor component of the Glasgow Coma Scale (mGCS), based on the area under the receiver operating characteristic (AUROC), a measure of discrimination. The tGCS is better able than the mGCS to discriminate patients with trauma who undergo neurosurgical intervention, have severe brain injury (traumatic brain injury [TBI]), or undergo emergency intubation from patients who do not experience these outcomes. Evidence on discrimination for identifying patients with severe injury (based on the Injury Severity Score [ISS] or criteria for trauma center need) was limited, but reported similar findings.[54,56] Although the differences in discrimination are statistically significant, their clinical significance is uncertain. The difference in the AUROC on each of these outcomes ranged from 0.03 to 0.05, or "small" based on our prespecified thresholds for interpreting differences in the AUROC. The tGCS was also better than the mGCS at discriminating trauma patients who died during hospitalization from those who survived hospitalization, but the difference in the AUROC was even smaller (0.01) than for non-mortality outcomes. Findings for the tGCS versus the Simplified Motor Scale (SMS) were similar to findings for the tGCS versus the mGCS for non-mortality outcomes, though the SMS performed slightly worse than the mGCS for mortality. Across scales, discrimination was generally higher for mortality (0.84 to 0.89) than for non-mortality outcomes (0.71 to 0.85). Although studies varied in how they defined neurosurgical interventions, severe brain injury, and emergency intubation, findings were generally similar across definitions for these outcomes. One study[61] in children found a greater difference in the AUROC (0.121, 95% confidence interval [CI] 0.068 to 0.174) compared with four studies conducted in mixed populations of adults and children (0.046, 95% CI 0.038 to 0.054), using a broader definition for severe brain injury that included interventions in addition to computed tomography (CT) imaging findings. However, the test for an interaction effect was not statistically significant (p=0.07).

Findings for discrimination were generally robust in sensitivity and subgroup analyses based on factors such as the age group analyzed (children vs. mixed populations of children and adults), study year (data collected after 2006 or included data collected before 2006), assessment setting (out-of-hospital vs. emergency department [ED]), or risk of bias ratings. However, sensitivity and subgroup analyses were limited by small numbers of studies, particularly for non-mortality outcomes. In addition, no study that evaluated mixed populations of adults and children reported results stratified by age group and no study that evaluated mixed populations of trauma patients reported results stratified by type of trauma. Therefore, findings are based on cross-study comparisons from stratified analyses. For age and type of trauma, few studies specifically evaluated children or patients with TBI, though those available reported findings similar to studies that evaluated mixed populations of adults and children or mixed trauma populations. None of the head-to-head studies specifically evaluated older patients or reported findings in this subgroup, though one study found lower accuracy of the tGCS using the standard cutoff score of 13 or less in adults older than 70 years of age versus those younger.[66] Another study found that among patients younger than 18 years of age, differences between the tGCS versus the mGCS on

discrimination for mortality and other outcomes were similar for children 0 to 3 years of age, in whom the verbal component is difficult to assess, and the whole cohort.[52] There were also no clear differences when studies were stratified according to out-of-hospital versus ED assessment of the Glasgow Coma Scale (GCS). In addition, studies that evaluated out-of-hospital versus ED scores in the same trauma center reported similar discriminative performance. There was insufficient evidence to determine how intubation status, intoxication status, receipt of field interventions, timing of GCS assessment, or level of training of people administering the GCS impacted predictive utility. No study evaluated how different GCS assessments performed in intoxicated patients or after intubation, or how performance varied according to receipt of out-of-hospital interventions. In the case of intoxication or intubation, the tGCS is often limited to the motor component due to the inability to accurately assess the verbal and eye domains. Studies on the effects of alcohol intoxication have shown somewhat mixed results, with some finding little effect on tGCS scores and others reporting lower tGCS scores in certain subgroups.[78-80] No study evaluated effects of the type or level of training of GCS assessors on predictive utility.

Several studies on discrimination for mortality utilized data from the National Trauma Data Bank (NTDB). Over 700 centers across the United States contribute to the NTDB.[81] This could result in double counting of patients analyzed in studies based on single centers and overweighting of such patients, if that center contributes data to the NTDB and depending on whether the NTDB and single center studies utilized data from the same time frame. There was insufficient information on the NTDB website to determine the extent to which trauma centers reported in single center studies contributed to NTDB. However, excluding NTDB studies had little impact on our findings regarding mortality, and estimates from the NTDB studies were very similar to the estimates from the studies conducted at single trauma centers. In cases where there were multiple studies from the same center (e.g., Loma Linda[20,36] and Denver)[48,52,60] and potential overlap in patient populations, restricting the analysis to the most recent study based on out-of-hospital GCS data also had little impact on findings.

Few studies reported the comparative diagnostic accuracy (sensitivity, specificity) of the tGCS versus the mGCS, but findings were generally consistent with analyses based on the AUROC. Based on standard cutoffs for the tGCS (≤13), mGCS (≤5), and SMS (≤1), differences in sensitivity and specificity were small. The consistency of findings between measures of discrimination and diagnostic accuracy may be expected, given that discrimination is calculated from sensitivity and specificity over a ranged of test cutoffs. There was insufficient evidence to compare the performance of triage instruments based on other measures of predictive utility, such as calibration or adjusted risk estimates.

No study evaluated how using the tGCS versus the mGCS or SMS impacts the likelihood of over- or under-triage. Head-to-head studies of the tGCS versus the mGCS or SMS were not designed to assess effects on over- or under-triage because the mGCS and SMS were taken from the tGCS, with each patient only undergoing a single triage decision. Studies that evaluate the tGCS, mGCS, or SMS alone are not helpful for assessing effects on over- or under-triage because they cannot isolate the effects of the GCS assessment from the many other factors that impact triage decisions.[4,82] Measures of diagnostic accuracy may provide some indirect indication of the potential degree of over- and under-triage, with 1-sensitivity indicating the proportion of patients who experience the outcome who would be missed (under-triage) and 1-specificity indicating the proportion of patients without the diagnosis (over-triage). However, this is an oversimplification that assumes that GCS assessments are the primary or sole driver of triage decisions, even though such decisions are known to be multifactorial and depend on other

patient factors (e.g., presence of hypotension, type of injury) and other variables (e.g., proximity to a trauma center). Nonetheless, as noted above, limited evidence suggests no marked differences in diagnostic accuracy between the tGCS, mGCS, and SMS, including two studies that attempted to assess overall impact on over- or under-triage based on diagnostic accuracy estimates.[18,54]

No study evaluated how using the tGCS versus the mGCS or SMS impacts the likelihood of clinical outcomes such as mortality, morbidity, or quality of life. As for over- or under-triage, understanding comparative effects on clinical outcomes requires head-to-head studies in which trauma patients who are assessed using different GCS scales are followed over time, in order to assess effects of the GCS scales versus other factors that impact clinical outcomes.

Evidence on interrater reliability and ease of use was limited. For assessment of patients with trauma, there was insufficient evidence to determine comparative interrater reliability of the tGCS, mGCS, and SMS, as there was only one head-to-head study with methodological limitations and imprecise estimates.[61] Other studies found the mGCS associated with higher interrater reliability than the tGCS, but were excluded because they did not report results separately for trauma patients;[20,36] other studies on interrater reliability were excluded because they focused on interrater reliability of the GCS among hospitalized patients. Studies that addressed ease of use were limited to those that evaluated whether the measures were scored correctly compared with a reference standard (usually expert assessment). Three studies found that the percentage of correct scores was higher for the mGCS than the tGCS,[71-73] though the difference was statistically significant in only one study.[71] Limited evidence suggests that errors are more frequent when assessing patient scenarios indicating moderate injury severity (tGCS scores of 9-13).[71,72,75] For both scales, use of a scoring aid or training appears to improve the proportion of correct scores. No study evaluated other measures of ease of use, such as time to complete the assessment or assessor satisfaction.

One study found that agreement between field and ED scores was similar for the tGCS and mGCS.[74] Although differences between field and ED scores were noted for both scales, the study also found that blood pressure readings changed. Therefore, some differences between field and ED scores may accurately reflect changing status of the patient due to receipt of out-of-hospital interventions and evolving clinical status, rather than true lack of agreement.

Findings in Relationship to What is Already Known

A prior Centers for Disease Control and Prevention (CDC) guideline found limited evidence on the predictive utility of the tGCS versus the mGCS.[4] Our review included a number of studies that were published after the CDC guideline and provides more robust findings, particularly since a number of studies evaluated very large samples, enabling precise estimates. Our findings on predictive utility are consistent with a prior systematic review comparing the tGCS versus the SMS that also found that discrimination was similar for these scales.[15] Like the CDC guideline, we found insufficient evidence to determine effects on over- or under-triage, or on clinical outcomes. Our finding that the interrater reliability of the tGCS and the mGCS are similar was based on very limited evidence in trauma patients; studies that included non-trauma patients have found the tGCS to be associated with lower interrater reliability.[23,83]

Applicability

Our findings on predictive utility of different GCS scales appear to have broad applicability to field triage in the United States, as they are based on large studies conducted in U.S. trauma

settings in mixed populations of adults and children with various types of trauma. We also restricted study inclusion to studies published after 1995, with most studies utilizing data collected through the last 5 to 10 years, suggesting high applicability to use in the context of current trauma systems.

Nonetheless, we identified a number of factors that can impact applicability. Despite the broad applicability of the evidence, its applicability to specific patient populations (e.g., specific type of trauma, age, presence and severity of intoxication, presence of medical comorbidities, and presence of other injuries) is less certain. For example, a modified version of the tGCS is utilized in young children[84] and the GCS was originally developed for assessment of TBI,[11] not trauma injuries in general. Limited evidence from across-study comparisons suggests similar results in children versus mixed populations of adults plus children and in patients with TBI versus mixed trauma populations. Within the subgroup of patients with TBI, the nature and prognosis of a TBI sustained from an impact injury (blunt force which may or may not involve fracture or intracranial lesion) may be different from that of a TBI sustained from an acceleration/deceleration injury[85] (diffuse injury resulting from contrecoup forces), and TBI often occurs in conjunction with other injuries. However, no study evaluated how the type of TBI injury or co-occurring injuries impacts performance of GCS scales. No study evaluated how predictive utility varied according to the level or training of field training personnel (e.g., Emergency Medical Technician [EMT], EMT-Intermediate, Advanced EMT/Paramedic, physicians, and nurses[25]). In fact, no study that used out-of-hospital scores reported the training of the people administering the GCS. Another factor that could impact applicability is that the performance of the tGCS and mGCS may be different when administered soon after injury (in the field) as opposed to later (after field stabilization and destination decisions have been made and patients have arrived in the ED). A number of studies on predictive utility were conducted in ED settings, which are more controlled and easier to study than field settings, [26] but may be of limited applicability to field settings. However, we found that predictive utility was similar in studies that utilized out-of-hospital versus ED GCS scores. We also found no clear differences in estimates of predictive utility when we restricted analyses to studies conducted in U.S. settings or to more recent (post 2006) studies, which may be more applicable to current U.S. practice.

The differences between the tGCS versus the mGCS or SMS in mean AUROC ranged from 0.01 to 0.05. This indicates that the ability of the scales to distinguish patients who experience an outcome from those who do not based on a higher score is 1 percent to 5 percent higher with the tGCS than with the more abbreviated scales. These differences were statistically significant, in part due to the large sample sizes evaluated in the studies. Although we classified such differences as "small," based on a priori thresholds, such thresholds are by nature somewhat arbitrary. The importance of "small" differences in discrimination depend in part on the seriousness of the outcome evaluated, the degree to which triage and other treatment decisions are based on the field triage scale, and the degree to which such actions impact clinical outcomes.

Studies on ease of use focused on the proportion of correct scores using video or written scenarios, as opposed to assessment of actual patients/situations. While some studies limited the time allowed for the assessment and utilized scenarios indicating a spectrum of injury intensity, the applicability of such studies to actual field conditions is uncertain. In addition, there was insufficient evidence to determine the effects of incorrect scoring on triage decisions (e.g., whether the incorrect score would result in a change from being above the threshold for transport to a high level trauma center to below, or vice versa).

Implications for Clinical and Policy Decisionmaking

Our review has implications for clinical and policy decisionmaking. Because we found no evidence on effects on clinical outcomes or risk of over- or under-triage, decisions regarding the selection of field assessment scales for trauma must rely on comparative predictive utility. Therefore, decisions regarding implementation of simplified GCS scales for field triage must consider the potential trade-offs between predictive utility and ease of use. Although the tGCS appears to have slightly greater discrimination than the mGCS or SMS for mortality, severe brain injury, and markers of severe injury such as receipt of neurosurgical interventions or emergency intubation, differences were relatively small and of uncertain clinical significance. It is also possible that some of the differences in predictive utility could be reduced because field triage personnel also use other factors to inform triage decisions. Limited evidence suggests that the mGCS may be easier to score correctly than the tGCS, which may offset disadvantages related to slightly lower predictive utility.[17,72,73]

Similar results for the mGCS and the SMS might be expected because the SMS utilizes the same information as the mGCS, with the only difference that patients with low scores on the mGCS (0 to 4) are collapsed into a single category for the SMS. Therefore, any differences in predictive utility between the SMS and mGCS are likely to be primarily related to ease of use and reliability. Therefore, studies in which the mGCS and SMS are derived from a single assessment are inadequate for evaluating comparative performance; rather, studies that independently apply the SMS and mGCS are needed.

Evidence on how factors related to patients, assessors, and settings impacts predictive utility is limited. However, even if such differences exist, there may be advantages to having a single scale that can be applied across trauma scenarios, instead of requiring field assessors to select from among different scales for particular situations, even if the predictive utility of the single scale is slightly lower in certain situations.

Although evidence on comparative interrater reliability was very limited, scales with poor interrater reliability would also be expected to be associated with low predictive utility, which depends in part on the reliability of assessments. As differences in predictive utility were small, differences in interrater reliability are unlikely to be a major factor driving clinical and policy decisions regarding selection of field assessment scales. For all field assessment scales, training and use of scoring aids is likely to improve reliability and accuracy of scoring.

Limitations of the Review Process

Our review process had some limitations. Because of anticipated heterogeneity due to differences in patient populations, outcomes, assessment settings, and other factors, we used the random effects DerSimonian-Laird model to pool data. Statistical heterogeneity was moderate or high in some analyses. The DerSimonian-Laird estimator can result in confidence intervals that are too narrow when statistical heterogeneity is present.[46] Therefore, we also performed analyses using an alternative random effects model, the profile likelihood method, to evaluate whether findings were sensitive to the random effects model used. Results were similar using the profile likelihood method. Even when statistical heterogeneity was high, estimates for differences in AUROC across studies were generally quite similar, with statistical heterogeneity largely related to the presence of large sample sizes and very precise estimates. For example, for the analysis of tGCS versus mGCS for mortality, the I^2 value was 0 percent in studies of adults and 0 percent in

studies of children, but the overall I^2 was 59 percent, based on a difference in pooled estimates of only about 0.01 (0.017, 95% CI 0.015 to 0.020 and 0.006, 95% CI 0.002 to 0.011, respectively).

Another limitation is that we had to impute CIs for some studies included in the pooled analyses. However, findings were similar when we used alternative imputation methods. Also, no head-to-head study performed the tGCS and mGCS or SMS separately. Rather, the mGCS and SMS were retrospectively determined for each patient from the tGCS. Therefore, the tGCS and mGCS or SMS were not performed independently, and it is uncertain how findings on the other GCS components may have impacted scoring on the motor component. Because data to estimate the correlation between tGCS and mGCS scores were limited, we assumed moderate correlation. Findings were similar in sensitivity analyses that utilized alternative correlation assumptions.

Another limitation is that most studies on predictive utility had methodological limitations, including failure to report attrition, missing data, and unclear methods for measuring outcomes. However, restricting analyses to low risk of bias studies had little impact on findings. Many of the studies of the tGCS and mGCS scale characteristics had major flaws and without including high and moderate risk of bias studies there would be no results to report, which is reflected in the insufficient and low rates of the strength of evidence.

We restricted analyses to English-language studies, which could result in language bias. However, we identified no foreign-language study that appeared to meet inclusion criteria, and our focus was on studies conducted in U.S. trauma settings, which are unlikely to be published in languages other than English. We were limited in our ability to assess publication bias, given the relatively small number of studies. Although we searched for relevant studies on ClinicalTrials.gov. This database focuses on clinical trials and did not identify any studies on predictive utility of the GCS, or relevant clinical trials.

As mentioned above, a number of studies were based on the large NTDB, which incorporates data from over 700 trauma centers across the United States, with very large sample samples. We could not reliably determine the degree to which studies that analyzed data from single centers analyzed populations with overlap with the NTDB studies. We performed sensitivity analyses in which NTDB studies were excluded, which had little impact on findings. Similarly, in situations where there was more than one study from a single trauma center with potential overlap, we found that results were similar when we focused on the most recent study from each center that used out-of-hospital GCS scores.

Gaps in the Evidence Base

The most important gap in the evidence base was the lack of evidence on effects of using different GCS scales on risk of over- or under-triage or on clinical outcomes. Although there were a fair number of head-to-head studies for predictive utility, including studies with large samples, the studies mostly evaluated mixed populations of adults and children with various types of trauma. Evidence to determine how predictive utility differs in subgroups defined by patient characteristics such as type of trauma (including presence and type of TBI), degree of intoxication, intubation status, and receipt of co-interventions was limited. There was also insufficient evidence to determine how the type or level of training of field personnel impacts predictive utility. A number of analyses and subgroup analyses were based on small numbers of studies, and should be interpreted with caution. We also identified little evidence on measures of predictive utility other than discrimination (e.g., diagnostic accuracy, calibration, adjusted risk estimates, and risk reclassification rates).

The literature on interrater reliability and ease of use was very limited. There was only one head-to-head study of interrater reliability in trauma patients with methodological limitations and imprecise estimates.[61] Studies on ease of use focused on scoring of written or video patient scenarios and did not address factors such as the time needed to complete the assessment or assessor satisfaction.

Future Research Needs

We identified several important future research needs. Head-to-head studies that assess one set of patients with the tGCS and another set with the SMS or mGCS are needed to understand effects on clinical outcomes as well as risk of over- or under-triage. Alternatively, studies that utilize the tGCS and simplified scales in the same patients could assess potential effects on over- or under-triage when the field assessment scales are incorporated into field assessment guidelines. However, such studies would represent less direct evidence, since they would not be based on actual differences in triage decisions. For over- and under-triage, studies should utilize standardized, validated measures. For predictive utility, prospective studies that independently assess patients using the tGCS and the mGCS or SMS would be useful for confirming the findings of the currently available retrospective studies, in which the mGCS or SMS were not independently assessed. Studies are needed to better understand the predictive utility in important subpopulations, including children, older patients, patients with specific types of trauma, and patients who have received field interventions prior to assessment. For patients who are intoxicated or intubated, studies that measure how frequently the tGCS reverts to the mGCS due to the inability to assess the other GCS components would be helpful. Studies that evaluate how the predictive utility of the tGCS compares with the mGCS or SMS according to the level of training of assessing personnel in the field are also needed. Studies that assess measures of predictive utility other than discrimination (e.g., calibration, adjusted risk estimates, diagnostic accuracy, risk reclassification) would be useful for providing more complete information on predictive utility. Finally, head-to-head studies on interrater reliability and ease of use (including time to use and assessor comfort or satisfaction) that are conducted on trauma patients in field settings are needed to better understand how these factors may impact decisions regarding selection of field assessment scales.

Conclusions

The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use. Research is needed to understand how use of the tGCS versus the mGCS or SMS impacts clinical outcomes and risk of over- or under-triage.

References

1. Centers for Disease Control and Prevention. 10 Leading Causes of Death by Age Group, United States - 2014. http://www.cdc.gov/injury/images/lc-charts/leading_causes_of_death_age_group_2014_1050w760h.gif. Accessed on May 5, 2016.
2. Centers for Disease Control and Prevention. 10 Leading Causes of Injury Deaths by Age Group Highlighting Unintentional Injury Deaths, United States - 2014. http://www.cdc.gov/injury/images/lc-charts/leading_causes_of_injury_deaths_unintentional_injury_2014_1040w740h.gif. Accessed on May 5, 2016.
3. Centers for Disease Control and Prevention. National Hospital Ambulatory Medical Care Survey: 2011 Emergency Department Summary Tables. http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf. Accessed on May 5, 2016.
4. Sasser SM, Hunt RC, Faul M, et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. MMWR Recomm Rep. 2012 Jan 13;61(RR-1):1-20. PMID: 22237112.
5. Faul M, Xu L, Wald MM, et al. Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002–2006. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control: 2010. http://www.cdc.gov/traumaticbraininjury/pdf/blue_book.pdf.
6. Centers for Disease Control and Prevention. Rates of TBI-related emergency department visits, hospitalizations, and deaths-United States, 2001–2010. Injury Prevention & Control: Traumatic Brain Injury. 2016.
7. Committee on Trauma, American College of Surgeons. Resources for Optimal Care of the Injured Patient. 2014. https://www.facs.org/quality%20programs/trauma/vrc/resources. Accessed June 20, 2016.
8. Baxt WG, Moody P. The impact of advanced prehospital emergency care on the mortality of severely brain-injured patients. J Trauma. 1987 Apr;27(4):365-9. PMID: 3573084.
9. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. N Engl J Med. 2006 Jan 26;354(4):366-78. doi: 10.1056/NEJMsa052049. PMID: 16436768.
10. Cook CH, Muscarella P, Praba AC, et al. Reducing overtriage without compromising outcomes in trauma patients. Arch Surg. 2001 Jul;136(7):752-6. PMID: 11448384.
11. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet. 1974 Jul 13;2(7872):81-4. PMID: 4136544.
12. Teasdale G, Maas A, Lecky F, et al. The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol. 2014 Aug;13(8):844-54. doi: 10.1016/S1474-4422(14)70120-6. PMID: 25030516.
13. Marshall LF. Evoked potentials: a decade later. Crit Care Med. 1991 Nov;19(11):1337. PMID: 1935149.
14. Bradshaw D. Report to the Surgeon General: Traumatic Brain Injury Task Force. May 15, 2007. 2008. https://www.hsdl.org/?view&did=482727.
15. Singh B, Murad MH, Prokop LJ, et al. Meta-analysis of Glasgow coma scale and simplified motor score in predicting traumatic brain injury outcomes. Brain Inj. 2013;27(3):293-300. doi: 10.3109/02699052.2012.743182. PMID: 23252405.
16. Brown KM, Hirshon JM, Alcorta R, et al. The implementation and evaluation of an evidence-based statewide prehospital pain management protocol developed using the national prehospital evidence-based guideline model process for emergency medical services. Prehosp Emerg Care. 2014;18 Suppl 1:45-51. doi: 10.3109/10903127.2013.831510. PMID: 24134543.

17. Beskind DL, Stolz U, Gross A, et al. A comparison of the prehospital motor component of the Glasgow coma scale (mGCS) to the prehospital total GCS (tGCS) as a prehospital risk adjustment measure for trauma patients. Prehosp Emerg Care. 2014 Jan-Mar;18(1):68-75. doi: 10.3109/10903127.2013.844870. PMID: 24329032.

18. Healey C, Osler TM, Rogers FB, et al. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. J Trauma. 2003 Apr;54(4):671-8; discussion 8-80. doi: 10.1097/01.ta.0000058130.30490.5d. PMID: 12707528.

19. Ross SE, Leipold C, Terregino C, et al. Efficacy of the motor component of the Glasgow Coma Scale in trauma triage. J Trauma. 1998 Jul;45(1):42-4. PMID: 9680010.

20. Gill M, Windemuth R, Steele R, et al. A comparison of the Glasgow Coma Scale score to simplified alternative scores for the prediction of traumatic brain injury outcomes. Annals of Emergency Medicine. 2005 Jan;45(1):37-42. PMID: 15635308.

21. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem. 2008 Jan;54(1):17-23. doi: 10.1373/clinchem.2007.096529. PMID: 18024533.

22. Menegazzi JJ, Davis EA, Sucov AN, et al. Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. J Trauma. 1993 Jan;34(1):46-8. PMID: 8437195.

23. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. Ann Emerg Med. 2004 Feb;43(2):215-23. doi: 10.1016/S019606440300814X. PMID: 14747811.

24. Neville AL, Nemtsev D, Manasrah R, et al. Mortality risk stratification in elderly trauma patients based on initial arterial lactate and base deficit levels. American Surgeon. 2011;77(10):1337-41 5p. PMID: 104607567.

25. National Emergency Medical Services Workforce Data Definitions. (Prepared by the American Institutes for Research through contract DTNH22-080F-00122.) National Highway Traffic Safety Administration. Washington, DC: February 2013. www.nhtsa.gov/staticfiles/nti/pdf/811720.pdf.

26. Badjatia N, Carney N, Crocco TJ, et al. Guidelines for prehospital management of traumatic brain injury 2nd edition. Prehosp Emerg Care. 2008;12 Suppl 1:S1-52. doi: 10.1080/10903120701732052. PMID: 18203044.

27. Newgard CD, Hsia RY, Mann NC, et al. The trade-offs in field trauma triage: a multiregion assessment of accuracy metrics and volume shifts associated with different triage strategies. J Trauma Acute Care Surg. 2013 May;74(5):1298-306; discussion 306. doi: 10.1097/TA.0b013e31828b7848. PMID: 23609282.

28. Haas B, Stukel TA, Gomez D, et al. The mortality benefit of direct trauma center transport in a regional trauma system: A population-based analysis. Journal of Trauma & Acute Care Surgery. 2012;72(6):1510-7 8p. PMID: 108124141.

29. Haas B, Gomez D, Zagorski B, et al. Survival of the fittest: the hidden cost of undertriage of major trauma. Journal of the American College of Surgeons. 2010;211(6):804-11 8p. doi: 10.1016/j.jamcollsurg.2010.08.014. PMID: 21036070.

30. Newgard CD, Staudenmayer K, Hsia RY, et al. The cost of overtriage: more than one-third of low-risk injured patients were taken to major trauma centers. Health Aff (Millwood). 2013 Sep;32(9):1591-9. doi: 10.1377/hlthaff.2012.1142. PMID: 24019364.

31. Newgard CD, Yang Z, Nishijima D, et al. Cost-Effectiveness of Field Trauma Triage among Injured Adults Served by Emergency Medical Services. J Am Coll Surg. 2016 Mar 3doi: 10.1016/j.jamcollsurg.2016.02.014. PMID: 27178369.

32. Zocchi MS, Hsia RY, Carr BG, et al. Comparison of Mortality and Costs at Trauma and Nontrauma Centers for Minor and Moderately Severe Injuries in California. Ann Emerg Med. 2016 Jan;67(1):56-67.e5. doi: 10.1016/j.annemergmed.2015.04.021. PMID: 26014435.

33. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2014. www.effectivehealthcare.ahrq.gov.

34. Graham E, Chou R, Newgard C, et al. Field triage guideline revision: Glasgow Coma Scale systematic review. PROSPERO 2016. CRD42016035944 Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016035944.

35. Bazzoli GJ, Madura KJ, Cooper GF, et al. Progress in the development of trauma systems in the United States. Results of a national survey. Jama. 1995 Feb 1;273(5):395-401. PMID: 7823385.

36. Gill M, Steele R, Windemuth R, et al. A comparison of five simplified scales to the out-of-hospital Glasgow Coma Scale for the prediction of traumatic brain injury outcomes. Academic Emergency Medicine. 2006 Sep;13(9):968-73. PMID: 16894005.

37. Baker SP, O'Neill B, Haddon W, Jr., et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma. 1974 Mar;14(3):187-96. PMID: 4814394.

38. Baxt WG, Jones G, Fortlage D. The trauma triage rule: a new, resource-based approach to the prehospital identification of major trauma victims. Ann Emerg Med. 1990 Dec;19(12):1401-6. PMID: 2240753.

39. Lerner EB, Willenbring BD, Pirrallo RG, et al. A consensus-based criterion standard for trauma center need. J Trauma Acute Care Surg. 2014 Apr;76(4):1157-63. doi: 10.1097/ta.0000000000000189. PMID: 24662885.

40. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013 Feb 19;158(4):280-6. doi: 10.7326/0003-4819-158-4-201302190-00009. PMID: 23420236.

41. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.

42. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1198-207. doi: 10.1016/j.jclinepi.2010.11.021. PMID: 21463926.

43. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993 Apr;39(4):561-77. PMID: 8472349.

44. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care. 2004 Dec;8(6):508-12. doi: 10.1186/cc3000. PMID: 15566624.

45. Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Ann Intern Med. 2013 Apr 2;158(7):544-54. doi: 10.7326/0003-4819-158-7-201304020-00006. PMID: 23546566.

46. Cornell JE, Mulrow CD, Localio R, et al. Random-effects meta-analysis of inconsistent effects: a time for change. Ann Intern Med. 2014 Feb 18;160(4):267-70. doi: 10.7326/m13-2886. PMID: 24727843.

47. Beskind DL, Keim SM, Spaite DW, et al. Risk adjustment measures and outcome measures for prehospital trauma research: recommendations from the emergency medical services outcomes project (EMSOP). Academic Emergency Medicine. 2011 Sep;18(9):988-1000. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01148.x. PMID: 21906205.

48. Thompson DO, Hurtado TR, Liao MM, et al. Validation of the Simplified Motor Score in the out-of-hospital setting for the prediction of outcomes after traumatic brain injury. Ann Emerg Med. 2011 Nov;58(5):417-25. doi: 10.1016/j.annemergmed.2011.05.033. PMID: 21803448.

49. Nance M, Rotondo M, Fildes J. National Trauma Data Bank 2013: Annual Report. 2013. https://www.facs.org/~/media/files/quality%20programs/trauma/ntdb/ntdb%20annual%20report%202013.ashx.

50. AHRQ Methods for Effective Health Care. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, eds. Methods Guide for Medical Test Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.

51. Agency for Healthcare Research and Quality. EPC Procedures Guide For Effective Health Care Program Evidence Reviews [Internal document]. Rockville, MD: Agency for Healthcare Research and Quality. Center for Evidence and Practice Improvement; 2014. http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318.

52. Acker SN, Ross JT, Partrick DA, et al. Glasgow motor scale alone is equivalent to Glasgow Coma Scale at identifying children at risk for serious traumatic brain injury. The Journal of Trauma and Acute Care Surgery. 2014 Aug;77(2):304-9. doi: http://dx.doi.org/10.1097/TA.0000000000000300. PMID: 25058258.

53. Al-Salamah MA, McDowell I, Stiell IG, et al. Initial emergency department trauma scores from the OPALS study: the case for the motor score in blunt trauma. Academic Emergency Medicine. 2004 Aug;11(8):834-42. PMID: 15289188.

54. Brown JB, Forsythe RM, Stassen NA, et al. Evidence-based improvement of the National Trauma Triage Protocol: The Glasgow Coma Scale versus Glasgow Coma Scale motor subscale. The Journal of Trauma and Acute Care Surgery. 2014 Jul;77(1):95-102; discussion 1-2. doi: http://dx.doi.org/10.1097/TA.0000000000000280. PMID: 24977762.

55. Caterino JM, Raubenolt A. The prehospital simplified motor score is as accurate as the prehospital Glasgow coma scale: analysis of a statewide trauma registry. Emergency Medicine Journal. 2012 Jun;29(6):492-6. doi: http://dx.doi.org/10.1136/emj.2010.110437. PMID: 21795294.

56. Cicero MX, Cross KP. Predictive value of initial Glasgow coma scale score in pediatric trauma patients. Pediatric Emergency Care. 2013 Jan;29(1):43-8. doi: http://dx.doi.org/10.1097/PEC.0b013e31827b52bf. PMID: 23283262.

57. Corrigan JD, Kreider S, Cuthbert J, et al. Components of traumatic brain injury severity indices. Journal of Neurotrauma. 2014 Jun;31(11):1000-7. doi: http://dx.doi.org/10.1089/neu.2013.3145. PMID: 24521197.

58. Davis DP, Serrano JA, Vilke GM, et al. The predictive value of field versus arrival Glasgow Coma Scale score and TRISS calculations in moderate-to-severe traumatic brain injury. J Trauma. 2006 May;60(5):985-90. PMID: 16688059.

59. Eken C, Kartal M, Bacanli A, et al. Comparison of the Full Outline of Unresponsiveness Score Coma Scale and the Glasgow Coma Scale in an emergency setting population. European Journal of Emergency Medicine. 2009 Feb;16(1):29-36. doi: http://dx.doi.org/10.1097/MEJ.0b013e32830346ab. PMID: 19106717.

60. Haukoos JS, Gill MR, Rabon RE, et al. Validation of the Simplified Motor Score for the prediction of brain injury outcomes after trauma. Annals of Emergency Medicine. 2007 Jul;50(1):18-24. PMID: 17113193.

61. Holmes JF, Palchak MJ, MacFarlane T, et al. Performance of the pediatric glasgow coma scale in children with blunt head trauma. Academic Emergency Medicine. 2005 Sep;12(9):814-9. PMID: 16141014.

62. Timmons SD, Bee T, Webb S, et al. Using the abbreviated injury severity and Glasgow Coma Scale scores to predict 2-week mortality after traumatic brain injury. J Trauma. 2011 Nov;71(5):1172-8. doi: http://dx.doi.org/10.1097/TA.0b013e31822b0f4b. PMID: 22071922.

63. Van de Voorde P, Sabbe M, Rizopoulos D, et al. Assessing the level of consciousness in children: a plea for the Glasgow Coma Motor subscore. Resuscitation. 2008 Feb;76(2):175-9. doi: 10.1016/j.resuscitation.2007.07.007. PMID: 17728046.

64. Nesiama JA, Pirallo RG, Lerner EB, et al. Does a prehospital Glasgow Coma Scale score predict pediatric outcomes? Pediatric Emergency Care. 2012 Oct;28(10):1027-32. doi: http://dx.doi.org/10.1097/PEC.0b013e31826cac31. PMID: 23023472.

65. Johnson DL, Krishnamurthy S. Send severely head-injured children to a pediatric trauma center. Pediatric Neurosurgery. 1996 Dec;25(6):309-14. PMID: 9348151.

66. Caterino JM, Raubenolt A, Cudnik MT. Modification of Glasgow Coma Scale criteria for injured elders. Academic Emergency Medicine. 2011 Oct;18(10):1014-21. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01164.x. PMID: 21951715.

67. Leijdesdorff HA, van Dijck JT, Krijnen P, et al. Injury pattern, hospital triage, and mortality of 1250 patients with severe traumatic brain injury caused by road traffic accidents. Journal of Neurotrauma. 2014 Mar 1;31(5):459-65. doi: http://dx.doi.org/10.1089/neu.2013.3111. PMID: 24093437.

68. Reisner A, Chen X, Kumar K, et al. Prehospital heart rate and blood pressure increase the positive predictive value of the Glasgow Coma Scale for high-mortality traumatic brain injury. Journal of Neurotrauma. 2014 May 15;31(10):906-13. doi: http://dx.doi.org/10.1089/neu.2013.3128. PMID: 24372334.

69. Majdan M, Steyerberg EW, Nieboer D, et al. Glasgow coma scale motor score and pupillary reaction to predict six-month mortality in patients with traumatic brain injury: comparison of field and admission assessment. Journal of Neurotrauma. 2015 Jan 15;32(2):101-8. doi: http://dx.doi.org/10.1089/neu.2014.3438. PMID: 25227136.

70. Takahashi C, Okudera H, Origasa H, et al. A simple and useful coma scale for patients with neurologic emergencies: the Emergency Coma Scale. American Journal of Emergency Medicine. 2011 Feb;29(2):196-202. doi: http://dx.doi.org/10.1016/j.ajem.2009.09.018. PMID: 20825789.

71. Bledsoe BE, Casey MJ, Feldman J, et al. Glasgow Coma Scale Scoring is Often Inaccurate. Prehospital & Disaster Medicine. 2015 Feb;30(1):46-53. doi: http://dx.doi.org/10.1017/S1049023X14001289. PMID: 25489727.

72. Feldman A, Hart KW, Lindsell CJ, et al. Randomized controlled trial of a scoring aid to improve glasgow coma scale scoring by emergency medical services providers. Annals of emergency medicine. 2015;65(3):325-9.e2. PMID: 25199613.

73. Heim C, Schoettker P, Gilliard N, et al. Knowledge of Glasgow coma scale by air-rescue physicians. Scand J Trauma Resusc Emerg Med. 2009;17:39. doi: 10.1186/1757-7241-17-39. PMID: 19723331.

74. Kerby JD, MacLennan PA, Burton JN, et al. Agreement between prehospital and emergency department glasgow coma scores. J Trauma. 2007 Nov;63(5):1026-31. PMID: 17993947.

75. Lane PL, Baez AA, Brabson T, et al. Effectiveness of a Glasgow Coma Scale instructional video for EMS providers. Prehospital & Disaster Medicine. 2002 Jul-Sep;17(3):142-6. PMID: 12627917.

76. Arbabi S, Jurkovich GJ, Wahl WL, et al. A comparison of prehospital and hospital data in trauma patients. J Trauma. 2004 May;56(5):1029-32. PMID: 15179242.

77. Dinh MM, Oliver M, Bein K, et al. Level of agreement between prehospital and emergency department vital signs in trauma patients. Emergency Medicine Australasia. 2013 Oct;25(5):457-63. doi: http://dx.doi.org/10.1111/1742-6723.12126. PMID: 24099376.

78. Rundhaug NP, Moen KG, Skandsen T, et al. Moderate and severe traumatic brain injury: effect of blood alcohol concentration on Glasgow Coma Scale score and relation to computed tomography findings. J Neurosurg. 2015 Jan;122(1):211-8. doi: 10.3171/2014.9.jns14322. PMID: 25361494.

79. Lange RT, Iverson GL, Brubacher JR, et al. Effect of blood alcohol level on Glasgow Coma Scale scores following traumatic brain injury. Brain Injury. 2010;24(7-8):919-27. doi: http://dx.doi.org/10.3109/02699052.2010.489794. PMID: 20545447.

80. Stuke L, Diaz-Arrastia R, Gentilello LM, et al. Effect of alcohol on Glasgow Coma Scale in head-injured patients. Ann Surg. 2007 Apr;245(4):651-5. doi: 10.1097/01.sla.0000250413.41265.d3. PMID: 17414616.

81. American College of Surgeons. National Trauma Data Bank 2015: Annual Report. 2015. https://www.facs.org/~/media/files/quality%20programs/trauma/ntdb/ntdb%20annual%20report%202015.ashx.

82. Newgard CD, Zive D, Holmes JF, et al. A multisite assessment of the American College of Surgeons Committee on Trauma field triage decision scheme for identifying seriously injured children and adults. J Am Coll Surg. 2011 Dec;213(6):709-21. doi: 10.1016/j.jamcollsurg.2011.09.012. PMID: 22107917.

83. Gill M, Martens K, Lynch EL, et al. Interrater reliability of 3 simplified neurologic scales applied to adults presenting to the emergency department with altered levels of consciousness. Annals of Emergency Medicine. 2007 Apr;49(4):403-7, 7.e1. PMID: 17141146.

84. Kirkham FJ, Newton CR, Whitehouse W. Paediatric coma scales. Dev Med Child Neurol. 2008 Apr;50(4):267-74. doi: 10.1111/j.1469-8749.2008.02042.x. PMID: 18312424.

85. Yokobori S, Bullock MR. Pathbiology of primary traumatic brain injury (Chapter 11). In: Zasler ND, Katz DI, RD Z, eds. Brain Injury Medicine: Principles and Practice. 2nd ed. New York: Demos Medical Publishing; 2013: p.137.

Abbreviations and Acronyms

Abbreviation	Definition
AEMT	Advanced Emergency Medical Technician
AHRQ	Agency for Healthcare Research and Quality
AIS	Abbreviated Injury Scale
AOR	Adjusted Odds Ratio
AUC, AUCROC, or AUROC	Area Under Receiver Operating Characteristic Curve
CCP	Critical Care Paramedic
CDC	Centers for Disease Control and Prevention
CER	Comparative Effectiveness Review
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CT	Computed Tomography
ED	Emergency Department
EMS	Emergency Medical Services
EMT	Emergency Medical Technician
GCS	Glasgow Coma Scale
HAPI	Health & Psychosocial Instruments
ICU	Intensive Care Unit
IQR	Interquartile Range
ISS	Injury Severity Score
KQ	Key Question
MEDLINE	Medical Literature Analysis and Retrieval System Online
mGCS	Motor Glasgow Coma Scale
n	Number
NHTSA	National Highway Traffic Safety Administration
NLR	Negative Likelihood Ratio
NPV	Negative Predictive Value
NR	Not Reported
NTDB	National Trauma Data Bank
NTTP	National Trauma Triage Protocol
OR	Odds Ratio
PENTA	Pediatric Trauma Registry
PICOTS	Populations, Interventions, Comparators, Outcomes, Timing, Types of Studies, and Setting
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Values
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
QUIPS	Quality in Prognostic Studies

Abbreviation	Definition
RCT	Randomized Controlled Trial
SD	Standard Deviation
SMS	Simplified Motor Score
SOE	Strength of Evidence
TBI	Traumatic Brain Injury
TEP	Technical Expert Panel
tGCS	Total Glasgow Coma Scale
TX	Texas
U.S. and USA	United States of America
vs.	Versus